BRÖSTCANCER FÖRBUNDET

Forskningsrapport

Huvudsökande:
Fredrik Strand
Överläkare, Docent
Karolinska Universitetssjukhuset

Frågeställning:
Kan artificiell intelligens användas för att identifiera kvinnor med ökad risk för att gå från mammografiscreening med oupptäckt cancer - och föreslå en annan undersökning i tillägg?

Tre frågor till Fredrik:

Hur kan resultatet av er forskning hjälpa patienterna, rent konkret?
Våra resultat visar att AI som analyserar mammografibilder kan hitta de som har hög risk trots en normal mammografi, och möjliggöra riktad kompletterande undersökning (t.ex. magnetkamera, MR, eller kontrastförstärkt mammografi). Det innebär att fler cancerfall kan upptäckas tidigare, de som missas av dagens mammografi, vilket i förlängningen kan förbättra prognosen och minska behovet av mer omfattande behandling. Problemet som funnits med tidigare metoder där urvalet till MR baserades på mammografisk täthet var att man behövder göra MR på 60 personer för att hitta en med cancer. Det blir ganska kostsamt. Med vårt AI-system hittar man en cancer per 15 personer. AI-modellen väljer alltså ut vilka som behöver MR på ett betydligt mer effektivt sätt.

Hur viktigt har stödet från Bröstcancerförbundet varit för er forskning?
Stödet har varit mycket viktigt och har bidragit till att vi kunnat genomföra både en utveckling av AI-system och en klinisk studie som lett till internationellt uppmärksammade resultat. Utan denna finansiering hade delar av arbetet inte kunnat genomföras i samma omfattning.

Vad vill du hälsa alla Bröstcancerförbundets givare?
Ert stöd gör verklig skillnad. Tack vare ert engagemang kan vi utveckla nya metoder som bidrar till att upptäcka bröstcancer tidigare och rädda liv.

Mer att läsa kring Fredriks forskning, samt publikationer finns på följande sidor.

Vetenskaplig Återrapportering

Fredrik Strand, Karolinska Universitetssjukhuset

Projektanslag 2022 & 2023 (F2023-0053)

Stockholm, 2026-04-13

Med stöd från Bröstcancerförbundet har vi kunnat delfinansiera forskning i vår grupp kring hur artificiell intelligens (AI) kan förbättra bröstcancerscreening. Arbetet har lett till viktiga resultat både från utveckling av nya AI-modeller och från en klinisk studie av dessa.

Vi har i en retrospektiv studie utvecklat en AI-modell som analyserar mammografibilder och kombinerar flera typer av information, såsom tidiga cancertecken, hur lätt cancer kan döljas i vävnaden, samt individuell risk. Denna modell visade sig vara betydligt bättre än dagens metoder för att identifiera kvinnor som senare utvecklar cancer.

Den största satsningen har varit den kliniska studien ScreenTrustMRI, som nyligen publicerats i Nature Medicine. I studien använder vi AI för att identifiera kvinnor som har en förhöjd risk för oupptäckt cancer trots normal mammografi, och erbjuder dessa kvinnor en kompletterande MR-undersökning.

Resultaten visar att metoden är mycket effektiv: genom att undersöka en liten grupp kvinnor (cirka 7%) kunde vi upptäcka många fler cancerfall än med tidigare metoder. Upptäckten var nästan fyra gånger mer effektiv än att använda traditionella mått på mammografisk täthet. Dessutom var de flesta tumörerna invasiva och kliniskt relevanta, vilket innebär att de sannolikt hade varit svåra att upptäcka i tid utan denna strategi.

Tillsammans visar dessa studier att AI kan bidra till att hitta cancer tidigare och mer träffsäkert, samtidigt som resurser som MR kan användas mer riktat. Detta är ett viktigt steg mot mer individanpassad bröstcancerscreening.

Arbetet fortsätter nu med att analysera långtidsutfall. Ett manuskript med de primära resultaten från 27-månadersuppföljningen i ScreenTrustMRI har idag skickats in för vetenskaplig publicering.

Jag tackar återigen för de generösa anslagen från Bröstcancerförbundet som möjliggjorde forskningen.


nature medicine

Article

https://doi.org/10.1038/s41591-

AI-based selection of individuals for supplemental MRI in population-based breast cancer screening: the randomized ScreenTrustMRI trial

nature medicine

Article

https://doi.org/10.1038/s41591-024-03093-5

AI-based selection of individuals for supplemental MRI in population-based breast cancer screening: the randomized ScreenTrustMRI trial

Received: 9 January 2024

Accepted: 23 May 2024

Published online: 8 July 2024

Check for updates

Mattie Salim[1,2], Yue Liu[3,4], Moein Sorkhei[3,4], Dimitra Ntoula[2], Theodoros Foukakis[1], Irma Fredriksson[5], Yanlu Wang[1], Martin Eklund[6], Hossein Azizpour[7], Kevin Smith[3,4] & Fredrik Strand[1,2]

Screening mammography reduces breast cancer mortality, but studies analyzing interval cancers diagnosed after negative screens have shown that many cancers are missed. Supplemental screening using magnetic resonance imaging (MRI) can reduce the number of missed cancers. However, as qualified MRI staff are lacking, the equipment is expensive to purchase and cost-effectiveness for screening may not be convincing, the utilization of MRI is currently limited. An effective method for triaging individuals to supplemental MRI screening is therefore needed. We conducted a randomized clinical trial, ScreenTrustMRI, using a recently developed artificial intelligence (AI) tool to score each mammogram. We offered trial participation to individuals with a negative screening mammogram and a high AI score (top 6.9%). Upon agreeing to participate, individuals were assigned randomly to one of two groups: those receiving supplemental MRI and those not receiving MRI. The primary endpoint of ScreenTrustMRI is advanced breast cancer defined as either interval cancer, invasive component larger than 15 mm or lymph node positive cancer, based on a 27-month follow-up time from the initial screening. Secondary endpoints, prespecified in the study protocol to be reported before the primary outcome, include cancer detected by supplemental MRI, which is the focus of the current paper. Compared with traditional breast density measures used in a previous clinical trial, the current AI method was nearly four times more efficient in terms of cancers detected per 1,000 MRI examinations (64 versus 16.5). Most additional cancers detected were invasive and several were multifocal, suggesting that their detection was timely. Altogether, our results show that using an AI-based score to select a small proportion (6.9%) of individuals for supplemental MRI after negative mammography detects many missed cancers, making the cost per cancer detected comparable with screening mammography. ClinicalTrials.gov registration: NCT04832594.

A full list of affiliations appears at the end of the paper. e-mail: fredrik.strand@ki.se

Mammography is the standard method for population-wide breast cancer screening, and randomized trials have confirmed that early cancer detection achieved through screening mammography leads to decreased breast cancer mortality[1,2]. However, around 30 percent of cancers in screened women are diagnosed as so-called interval cancers, becoming symptomatic after a negative screening and before the next planned screening[3]. Interval cancers may escape screening as they may be fast-growing and not present on mammograms at the time of screening, were missed by the radiologist reader or were undetectable by mammography at the time[4,5]. Like other symptomatically detected cancers, interval cancers show aggressive biology and have relatively poor prognosis[6,7].

Contrast-enhanced MRI has been shown to be a more sensitive method for early detection of breast cancer compared with mammography[8]. The sensitivity of mammography is reduced for women with extremely dense breasts, while the sensitivity of MRI is unaffected[9–11]. In addition, women with dense breast tissue experience up to a twofold increase in breast cancer risk compared with the average woman[12–17]. The randomized DENSE trial conducted in the Netherlands demonstrated a 50% reduction in interval cancer for women invited to MRI compared with those not invited, and more than 80% reduction for women actually undergoing MRI compared with those not invited[11]. Based on those results, the cost-effectiveness ratio of offering MRI every 3 years to women with extremely dense breasts (around 8–10% of the population) has been estimated at €37,181 per QALY[18]. Due to the relatively high cost, MRI has not yet been included in any national screening program despite the European Society of Breast Imaging (EUSOBI) 2022 recommendation that women aged 50–70 years with extremely high mammographic density should have supplemental imaging by MRI or, if MRI is not available, by other methods, at least every 4 years (preferably every 2–3 years)[19].

Previous studies have shown that reductions in the number of interval cancers can be achieved by using MRI for supplemental screening of women with high mammographic density[11,20]. The increased detection, however, comes at a substantially increased per examination cost, which results in lower cost-effectiveness compared with mammography and hinders widespread implementation[21,22]. Here, we report interim results of a prospective clinical study using a new AI image tool to select high-risk women for supplementary MRI.

The ScreenTrustMRI trial (NCT04832594) is a randomized controlled trial to assess the use of an AI tool—AISmartDensity—to select individuals for supplemental MRI. AISmartDensity was developed by the authors (Y.L., M. Sorkhei and K.S.) using data collected by authors M. Salim. and F.S. The AI tool has a modular structure with three component models assessing underlying risk, potential masking and suspicious cancer signs (Supplementary Information). The primary use case for the AISmartDensity algorithm is to triage women for supplemental MRI after negative screening mammography. The AI tool was evaluated in a retrospective clinical study where it showed a cancer detection rate up to 29 cancers per 1,000 examinations in an 8% selection of the population[23]. The primary endpoint of the ScreenTrustMRI trial is to examine the incidence of advanced cancer at 27-month follow-up after the initial screening in the group of individuals randomized to MRI compared with those randomized to no MRI. Advanced cancer, termed 'delayed' in the study protocol, was defined as any of the following: interval cancer, invasive component larger than 15 mm and lymph node positive cancer (that is, N1), or screen-detected cancers meeting specific criteria such as interval cancers, cancers having an invasive component larger than 15 mm or lymph node metastasis. Secondary endpoints include cancer detection at supplemental MRI, participant engagement, AI score distribution, tumor characteristics, radiological process measures and questionnaire responses. The trial started inclusions on 1 April 2021 and ended inclusions on 7 April 2023, and the follow-up period will end in August of 2025 for the last included patient. The current report focuses on the detection of additional cancers for individuals randomized to, and undergoing, supplemental MRI. It was prespecified in the study protocol that secondary endpoints could be reported before primary endpoints.

Our hypothesis was that AI-based image analysis will provide a more effective selection tool than traditional density in terms of the proportion of MRI examinations leading to a cancer diagnosis.

Results

Study population

The source screening population for the ScreenTrustMRI trial consisted of 59,354 women whose mammograms were screened with AISmartDensity, of whom 4,103 were eligible to participate due to a 'very high' (top 6.9%) AISmartDensity score. The distribution of BI-RADS (Breast Imaging-Reporting and Data System) breast density[24] and AISmartDensity assessments are shown in Supplementary Table 1. Due to an invitation error caused by delayed image processing, 282 women did not receive invitations, resulting in an invited population of 3,821 (Fig. 1).

Out of 1,315 (34% of 3,821) individuals who accepted the invitation, 663 (50% of 1,315) individuals were randomized to undergo MRI. Among those randomized to MRI, 104 (16% of 663) chose not to undergo the examination, while 559 (84% of 663) completed the MRI examination to form the final cohort for this analysis (Fig. 1). The median age of the cohort was 56 years (interquartile range (IQR) 50–65 years) (Table 1). Among participants, 22 (4% of 559) had a previous history of breast cancer and 104 (19% of 559) individuals reported having a family member with a history of breast cancer.

MRI reads and subsequent procedures

A summary of the MRI results and subsequent procedures is reported in Table 2, including the radiologist assessment of each MRI examination, which was read by two breast radiologists (F.S., D.N.) with an experience level of 4 and 5 years, respectively, in MRI reading. For each MRI examination, the radiologists assessed fibroglandular proportion, corresponding to mammographic density, as well as the background parenchymal enhancement. Any lesions found were classified according to BI-RADS to assess the degree of malignancy suspicion of imaging findings: 464 (83%) were BI-RADS 1–2 ('benign'); 54 (9.7%) were BI-RADS 3 ('probably benign'); 27 (4.8%) were BI-RADS 4 ('suspicious for malignancy') and 14 (2.5%) were BI-RADS 5 ('highly suggestive of malignancy'). In Sweden, it is common practice to biopsy BI-RADS 3 lesions and avoid short-term follow-up. Therefore, our study protocol included a first attempt at biopsy using traditional methods, but, if not possible, we applied a short-term follow-up protocol for women with BI-RADS 3 lesions. There were 95 (17%) examinations that met the condition of BI-RADS 3–5 and were flagged accordingly for further work-up. Diagnostic biopsies were obtained for 12.7% (71 of 559) of the study cohort; 71 of the 95; 62 (65%) by guided ultrasound and 9 (9.5%) by MRI. Of the 24 flagged cases who had no biopsy, 23 were BI-RADS 3 lesions that could not be localized on second-look ultrasound and 1 was BI-RADS 4 for which no corresponding lesion could be localized on second-look ultrasound or during an attempt at MRI-guided biopsy; these patients had MRI follow-up examinations scheduled at 6, 12 and 24 months per protocol; complete results for those are not yet available. In the analysis, the 24 BI-RADS 3–5 lesions that were not subjected to biopsy were classified as incomplete rather than benign, considering the unavailability of comprehensive results. Table 2 also provides an overview of the MRI assessments for individuals who were diagnosed with breast cancer and those who were not. Among the diagnosed individuals, the most common type of lesions were enhancing masses; however, among those without cancer, nonmass enhancements were more frequent.

Diagnostic outcomes

Cancerous lesions were detected in 36 participants, corresponding to 64.4 (95% confidence interval (CI) 46.8–88.1) cancer detection rate per

https://doi.org/10.1038/s41591-024-03093-5

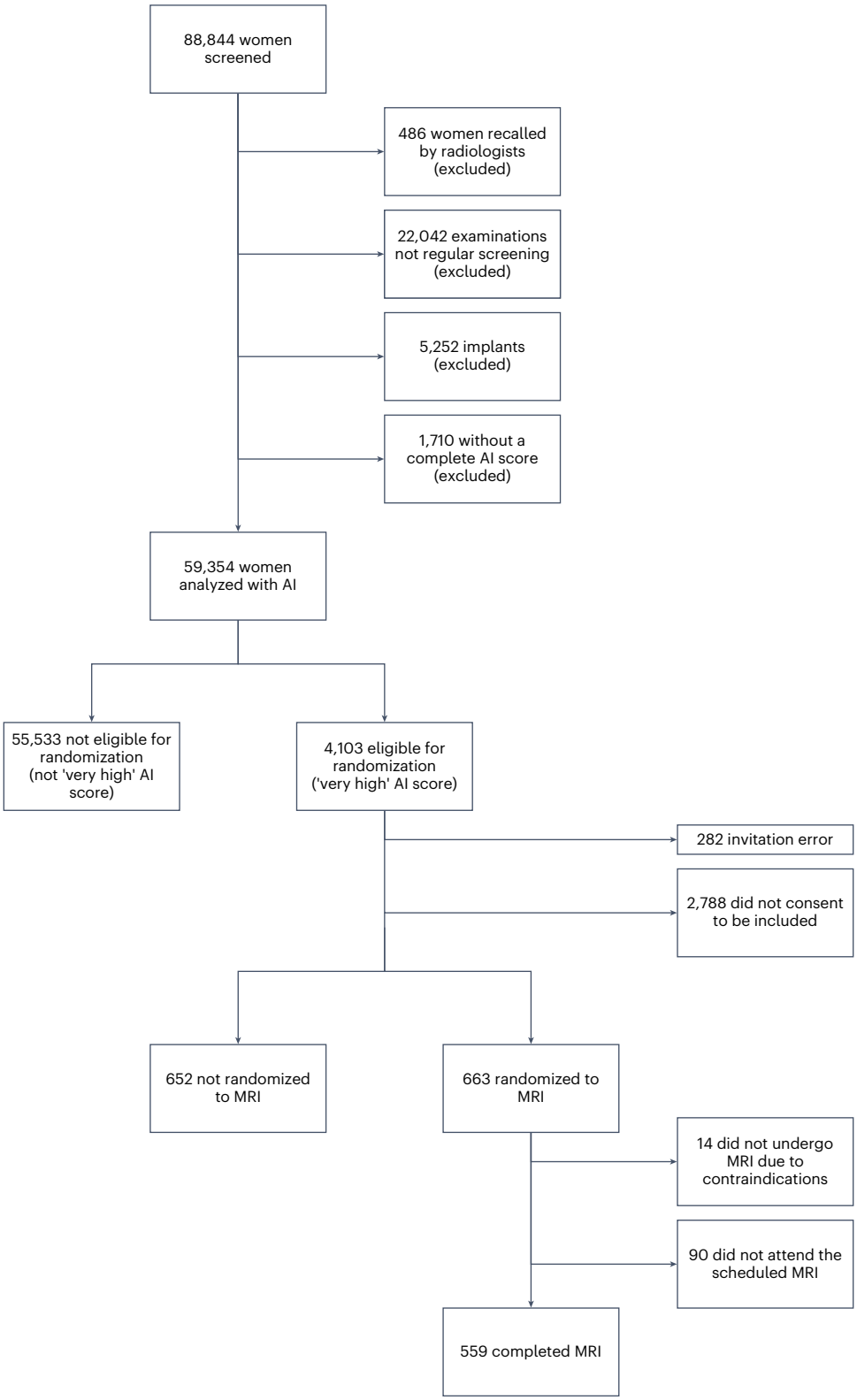

Fig. 1 | Study population. CONSORT diagram for the study population.

1,000 MRI examinations (Table 3). The proportion of cancers identified in individuals recalled after MRI (positive predictive value 1 (PPV1)) was 37.9% (95% CI 28.6 to 48.1). For individuals assessed as BI-RADS 3, 4 and 5, the PPV1 was 13.0%, 63.0% and 85.7% respectively. The proportion of cancers among biopsied individuals (positive predictive value 3 (PPV3)) was 50.7% (95% CI 39.0–62.3). Figure 2 displays four cases highlighting suspicious findings on MRI, enabling comparison with the preceding negative screening mammogram. Supplementary Fig. 1 shows the results normalized to a screening cohort of 10,000 individuals with negative mammography findings and Supplementary Table 2 presents the overall AISmartDensity score and the metrics for each of the three component models across all 36 cancers.

https://doi.org/10.1038/s41591-024-03093-5

Table 1 | Description of the study population of individuals undergoing MRI (*n*=559)

All	Value
Age, years (0 missing), median (IQR)	56 (50–65)
40–49 (*n*=129)	46 (43–47)
50–59 (*n*=214)	54 (52–56)
60–69 (*n*=145)	65 (63–67)
≥70 (*n*=71)	72 (71–73)
Weight, kg (24 missing) (median, IQR)	69 (63–77)
Height, cm (13 missing) (median, IQR)	167 (163–172)
Breast self-examination (21 missing) (*n*, %)	
Never	62 (11.5%)
Occasionally	400 (74%)
Often	76 (14.1%)
Previous breast MRI (20 missing) (*n*, %)	26 (4.7%)
Previous breast cancer (13 missing) (*n*, %)	22 (4.0%)
Previous breast disease (18 missing) (*n*, %)	64 (11.8%)
Breast cancer relative (12 missing) (*n*, %)	104 (19%)
Ovarian cancer relative (14 missing) (*n*, %)	22 (4.0%)
Age at menarche (39 missing) (*n*, %)	
11 years or younger	77 (14.8%)
12–14 years	375 (72.1%)
>14 years	68 (13.1%)
Number of pregnancies (29 missing) (n, %)	
0 pregnancies	30 (5.66%)
1–2 pregnancies	244 (46.0%)
≥3 pregnancies	256 (48.3%)
Age first pregnancy (77 missing) (*n*, %)	
<19 years	46 (9.5%)
20–29 years	269 (55.8%)
>30 years	167 (34.6%)
Number of births (44 missing) (*n*, %)	
1 birth	91 (17.7%)
2–3 births	369 (71.7%)
≥4 births	19 (3.69%)
History of breastfeeding (90 missing) (n, %)	
<1 month	24 (5.1%)
1–11 months	235 (50%)
≥12 months	210 (44.8%)

Percentages do not always sum to 100% due to rounding and some values are missing due to incomplete participant reporting.

Cancer characteristics

A detailed overview of the MRI and histopathological characteristics of the cancers detected is shown in Table 4. The median size of the only, or the largest, malignant lesion measured on MRI images was 12 mm (IQR 8–18 mm), which largely agreed with the histopathological analysis of the surgical specimens, which showed 13 mm (IQR 8–17). The total extent of malignancy ranged between 7 and 85 mm with a median of 15 mm (IQR 12–35). A total of 13 (36% of 36) were larger than 20 mm, corresponding to stage 2 or higher. Among all the diagnosed cancers, 7 (19% of 36) presented with multiple mass lesions on MRI whilst histopathological analysis confirmed multifocality for 4 (11% of 36). Only 3 (8% of 36) cases had lymph node metastases. In the histopathological analysis of the surgical specimens, most (22 or 61%) were a combination of invasive and ductal cancer in situ, with 5 (14% of 36) being in situ only. In terms of histological origin, according to the traditional classification system, 27 (75% of 36) cancers were classified as ductal breast cancer in the surgical specimens.

Discussion

In this report from the ScreenTrustMRI trial, we assessed the impact of applying the AISmartDensity tool to identify women at risk of undetected cancer following negative mammography screening, focusing on those who could potentially benefit from supplemental MRI. We observed a cancer detection rate of 64.4 cancers per 1,000 MRI examinations, and a positive predictive value of 38% for individuals recalled after MRI and 50.7% for individuals who were biopsied. Most of these malignant lesions had invasive components. The invasive cancers had a median size of 13 mm on pathology analysis, which is smaller than the average size of 15.8 mm and 19.6 mm for mammography screen-detected cancer and interval cancer, respectively, previously reported for a similar breast center in the Stockholm area[25]. Using the AISmartDensity method would make the detection cost per cancer similar to the cost in population-wide screening mammography and contribute to earlier detection of invasive cancer. The cancer detection rate of our trial at 64 (95% CI 46.8–88.1) cancers per 1,000 MRIs corresponds to about 3.8 times higher supplemental cancer detection rate compared with the traditional density method used in the DENSE trial[11] at 16.5 cancers per 1,000 MRIs. A cost-effectiveness study based on the results from the DENSE trial estimated the cost per QALY gained for supplemental MRI every 3 years at €37,181. Our results suggest that using AISmartDensity, the cost per QALY would probably be markedly lower given the close to four times higher supplemental cancer detection rate. As a alternative approach with potential to reduce MRI utilization and further increase cost-effectiveness, it may be suggested that radiologists should re-review the mammograms of all women with a very high AISmartDensity score. An assessment of this approach is planned as a post hoc reader study.

A recent meta-analysis[8] evaluated various supplementary screening methods in women with dense breasts and a negative screening mammogram across 22 studies and a total of 132,166 individuals. Of the 22 studies, 3 focused on supplementary screening with MRI[11,26,27], totaling 43,577 individuals screened, of whom 7,021 had dense breasts and were eligible for supplemental MRI. The pooled estimate of cancer detection rate for MRI was 25.7 per 1,000 MRI exams, with PPV1 27.7% and PPV3 34.1%. MRI was superior in terms of cancer detection compared with other supplemental screening methods: digital breast tomosynthesis, handheld ultrasound and automated breast ultrasound. One of the evaluated trials was the DENSE randomized trial, in which Bakker et al. assessed the effectiveness of supplemental MRI screening for individuals aged 50–75 years selected by traditional mammographic density[11]. Breast density was measured using automated software, and individuals with a BI-RADS D density score (corresponding to dense breast tissue) were invited to the study. The results showed that early detection by supplemental MRI screening led to a decrease in interval cancers by more than 80 percent. In the DENSE trial, the cancer detection rate among women undergoing MRI was 16.5 per 1,000 MRI examinations with a PPV for recall of 17.4%. The second trial[26] included females aged 40–70 years of average risk (lifetime risk of 6–12%) assessed using the Gail model that uses personal medical and reproductive history, as well as the breast cancer history among first-degree relatives, but does not include breast density as a risk factor[28]. The supplementary cancer detection rate in the study was 22.6 per 1,000 examinations. The third study[27] compared abbreviated MRI protocols (AP) with full diagnostic protocols (FP). The study included females aged 30–71 years with dense breasts (per BI-RADS) and normal mammography results, although the indication for supplementary examination in this study was unclear to us. The cancer detection rate

https://doi.org/10.1038/s41591-024-03093-5

Table 2 | MRI radiologist reads and procedures (*n*=559)

	Overall (*n*=559) *n* (%)		No breast cancer (*n*=523)[a] *n* (%)		Diagnosed breast cancer (*n*=36) *n* (%)		*P* value[b]
Amount of fibrograndular tissue							
Almost entirely fat	4	(0.7%)	4	(0.77%)	0	(0%)	N/A
Scattered fibroglandular tissue	206	(36.9%)	188	(36%)	18	(50%)	0.098
Heterogenous fibroglandular tissue	288	(52%)	275	(53%)	13	(36%)	0.057
Extreme fibroglandular tissue	59	(10.6%)	54	(10.3%)	5	(13.9%)	0.508
Background parenchymal enhancement							
Minimal	174	(31.3%)	161	(30.8%)	13	(36%)	0.541
Mild	230	(41.1%)	212	(40.5%)	18	(50%)	0.296
Moderate	114	(20.4%)	110	(21%)	4	(11%)	0.153
Marked	34	(6.1%)	33	(6.3%)	1	(2.8%)	0.397
BI-RADS abnormality score							
1	346	(61.9%) (464 (83%))	346	(66%)	0	(0%)	N/A
2	118	(21.1%)	118	(23%)	0	(0%)	N/A
3	54	(9.7%) (95 (17%))	47	(9%)	7	(19.4%)	0.046
4	27	(4.8%)	10	(1.9%)	17	(47.2%)	<0.001
5	14	(2.5%)	2	(0.38%)	12	(33.3%)	<0.001
Type of lesion (BI-RADS 3–5)							
Focus	11	(1.9%)	9	(15.3%)	2	(5.5%)	0.136
Nonmass enhancement	28	(5%)	23	(39%)	5	(13.9%)	0.006
Mass	50	(8.9%)	21	(36%)	29	(81%)	<0.001

[a]Includes women that were biopsied with a benign finding and women that are undergoing follow-up. [b]*P* value for the comparison between no breast cancer and diagnosed breast cancer. Percentages do not always sum to 100% due to rounding. Numbers do not always equal 559 due to missing radiologist assessments. N/A, not available.

Table 3 | Breast cancer detection by supplemental MRI

BI-RADS	Examinations, *n*	Breast cancer detected				No breast cancer detected		
		Cancer detection rate		PPV of BI-RADS 3–5 (%)		Not recalled	Benign biopsy	Incomplete MRI follow-up
		n	per 1,000 MRI exams	PPV1 (95% CI)	PPV3 (95% CI)			
All	559	36	64.4 (95% CI 46.8–88.1)			464		
3–5	95	36		37.9 (28.6–48.1)	50.7 (39.0–62.3)		35	24
3	54	7		13.0 (6.2–25.1)	22.6 (10.8–41.2)		24	23
4	27	17		63.0 (42.8–79.4)	65.4 (44.7–81.5)		9	1
5	14	12		85.7 (53.5–96.9)	85.7 (53.5–96.9)		2	0

Two women decided to not pursue MRI follow-up. PPV1, PPV of BI-RADS 3–5 read; PPV3, PPV of biopsy performed after BI-RADS 3–5 read; 95% CI, estimated 95% CI by Stata 'proportion' command.

was between 31 for AP and 33 for FP per 1,000 MRI examinations, while the PPV for recall varied between 27.8% (AP) and 41% (FP). It is notable that the additional cancer detection rate in the present study was two to four times higher than the three studies included in the meta-analysis. Traditional mammographic density and risk models thus seem to capture a markedly lower amount of relevant image information compared with the AISmartDensity tool.

When comparing mammographic density with AISmartDensity in the source population, we note associations between different mammographic density categories and the 'very high' category of AISmartDensity. Across all mammographic density categories, most individuals were classified as not 'very high.' However, within each density category there were notable variations in the proportions classified as 'very high.' Individuals categorized as BI-RADS D had the highest proportion 12% (682 of 5,515) classified as 'very high', followed by BI-RADS C 11% (3055 of 27742) and BI-RADS B 1.9% (358 of 19170). In contrast, those with BI-RADS A had the lowest proportion 0.1% (8 of 6927) classified as 'very high.' These findings indicate that there is some association between traditional mammographic density and AISmartDensity but far from full agreement.

Not all women with a very high density (BI-RADS D) by traditional measures are included in the very high AISmartDensity category. An important purpose of density classification is to ensure that we efficiently allocated supplemental screening resources to find as many undetected cancers as possible given a specific budget or cost-effectiveness goal. In that capacity, our results suggest that using AISmartDensity is superior to traditional density. It should be noted that, even if AISmartDensity is likely to catch around four times as many cancers as using BI-RADS D, there could be some cancers identified by BI-RADS D that AISmartDensity 'very high' might miss.

The results in terms of cancers detected should be put in perspective of what would be the expected future cancers without our intervention. Once the 27-month follow-up time is complete in our trial, we can compare with the women randomized to not have MRI. At present, an approximation is still possible. There were 559 individuals having MRI, which forms 6.9% of a total population of 8,101 individuals.

https://doi.org/10.1038/s41591-024-03093-5

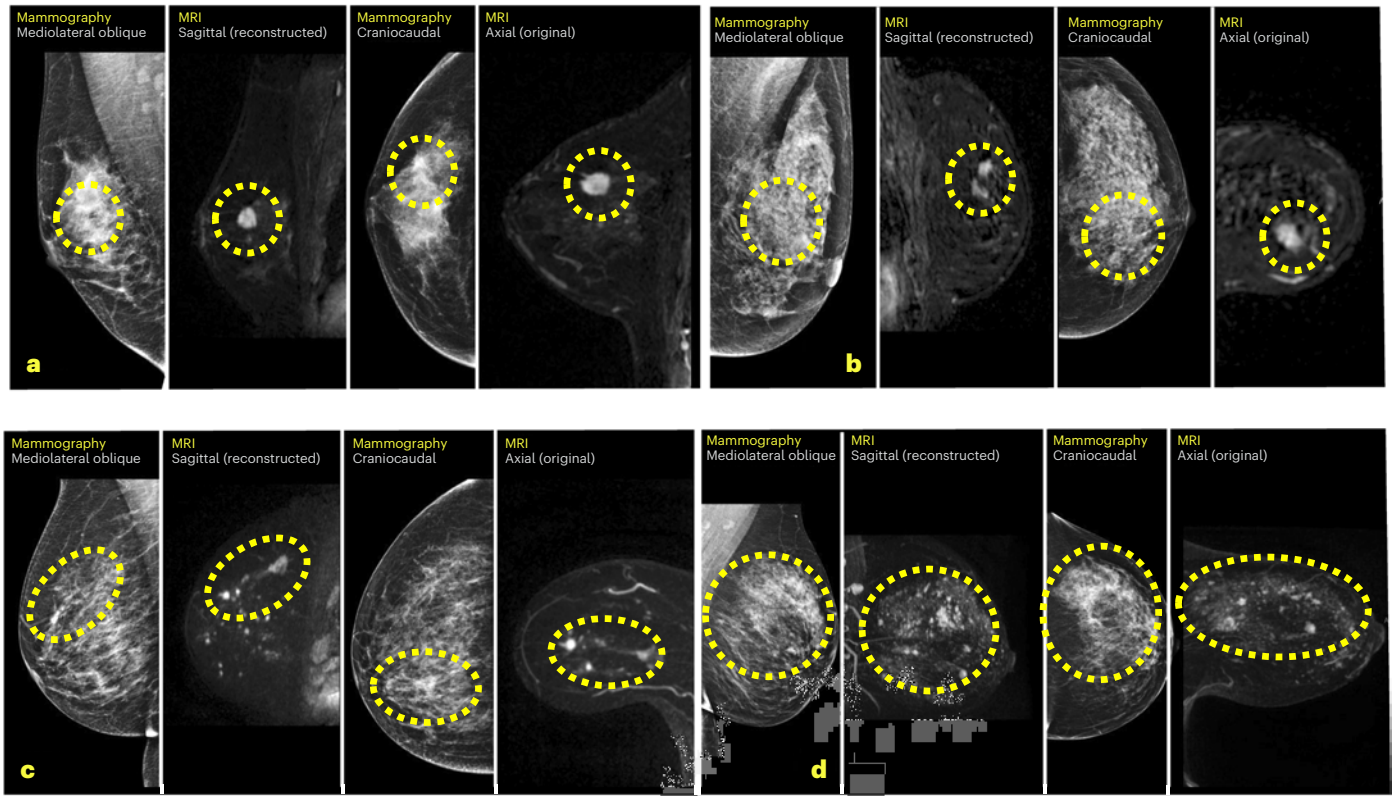

Fig. 2 | Cancers detected by supplemental MRI after negative mammography. **a**, Case A was a 13-mm large BI-RADS 4 finding on MRI that corresponded to a 13-mm large invasive cancer as diagnosed in the surgical specimen. **b**, Case B was a 9-mm large BI-RADS 4 finding on MRI that corresponded to an 8-mm invasive cancer and 14-mm ductal cancer in situ. **c**, Case C was three suspicious lesions within a total area of 60 mm, BI-RADS 5 finding on MRI, that corresponded to a 50-mm multifocal invasive lobular cancer. **d**, Case D was two suspicious lesions, the largest 13 mm, BI-RADS 3 on MRI, which corresponded to a 10-mm invasive lobular cancer with 85-mm extent including ductal cancer in situ.

The expected number of interval cancers in the underlying population would be 16 at the approximate 0.2% interval cancer rate in Sweden, and the expected number of next-round screen-detected cancers would be 41 at the approximate 0.5% cancer detection rate in the regular Swedish breast cancer screening system. The detected 36 cancers in the selected individuals correspond to 63% of the 57 expected future cancers in the entire screening population. The potential to pre-emptively detect most cancers by offering MRI to a small proportion of individuals represents an important healthcare value proposition.

In hindsight, some of the mammograms could be seen as showing subtle suspicious findings (Fig. 2). To understand whether this would be enough for radiologists to correctly recall them, all examinations included in this trial will be assessed in the planned post hoc reader study.

For mammography, examinations classified as BI-RADS 3 should ideally have a PPV below 2%. However, the interpretation of BI-RADS 3 in MRI is not as firmly established due to the lack of knowledge regarding features that constrain the risk to the 2% threshold. The relatively high PPV of 13% in the current study indicates that the radiologists involved were not fully calibrated to the expected 2% level and underscores the challenge in classifying BI-RADS 3 for MRI.

It is notable that most of the cancers detected in the population selected for supplementary MRI exhibited invasive features. This could be attributed to a combination of the AI tool and the use of MRI. It is known that MRI has increased sensitivity for invasive breast cancer, due primarily to increased contrast enhancement[29], compared with more indolent cancer in situ[30].

Lymph node metastases were relatively rare in our study. Only 8% of the 36 cases demonstrated lymph node metastasis, with 6% of cases showing evidence of this during the initial biopsy and 8% in the surgical specimen. These findings may be indicative of early detection when using MRI as a supplemental screening method. The larger tumor size observed in some of the lesions may be partially attributed to the prevalent screening nature of this study. We anticipate that the frequency of larger tumors will diminish in subsequent screening rounds. As per our study protocol, after 27 months follow-up we will report differences between the MRI arm and the control arm in terms of the number of cancers with adverse prognostic features.

Based on results from the previously published retrospective evaluation, it would seem that the cancer signs part of the AISmart-Density score would be just as good, in terms of AUC, as the summary score. However, in the current prospective implementation, 16 of 36 cancers actually had a higher masking or risk score than cancer signs score (Supplementary Table 2)[23]. The ground truth on which the AUC calculation is based naturally differ between the retrospective evaluation (based on a 3-year follow-up time) and the prospective implementation (cancers that near time of screening are actually detected by MRI). We believe that the three components adds robustness and are valuable in practice.

The main strength of this study is its prospective nature, using a population derived from a well-controlled prespecified randomized clinical trial conducted within a population-based screening program. The use of the automated AISmartDensity assessment ensured there was no interreader variability in mammogram assessment. The software was developed in-house (patent pending) and has not yet received CE mark or United States Food and Drug Administration approval. A key limitation of the present report is that the cancer detection rate using this method can be compared only with results using traditional density in other studies, not within this study. Also, a limitation is that we do not have patient parameters required to compare with traditional risk model scores. Furthermore, although this study provides clear

https://doi.org/10.1038/s41591-024-03093-5

Table 4 | Characteristics of cancers detected by supplemental MRI (*n*=36)

	Median (IQR) or *n* (%)
MRI	
Lesion diameter, mm, median (IQR)	12 (8–18)
Multiple lesions, *n* (%)	
No	29 (81%)
2 lesions	5 (14%)
≥3 lesions	2 (6%)
Biopsy (*n*=36)	
Invasiveness, *n* (%)	
In situ only	4 (11.1%)
Invasive only	27 (75%)
In situ and invasive	5 (13.9%)
Histology, *n* (%)	
Ductal	30 (83.3%)
Lobular	5 (13.9%)
Mucinous	1 (2.8%)
Grade, *n* (%) (1 missing)	
1	6 (17%)
2	26 (74.3%)
3	3 (8.6%)
Receptor expression, *n* (%)	
HR^+	33 (91.7%)
$HER2^+/HR^-$	0 (0%)
$HER2^-/HR^-$	3 (8.3%)
Ki-67, *n* (%) (5 missing)	
Low	5 (14%)
Intermediate	22 (61%)
High	4 (11%)
Lymph node, *n* (%)	
No biopsy	31 (86%)
Metastasis	2 (6%)
No metastasis	3 (8%)
Surgical specimen (*n*=35; 1 in neoadjuvant therapy)	
Lesion diameter, mm, median (IQR)	13 (8–17)
Total extent, mm, median (IQR)	15 (12–35)
Total extent, mm, min and max	7 and 85
Multiple lesions, *n* (%)	
No	32 (89%)
2 lesions	2 (6%)
≥3 lesions	2 (6%)
Invasiveness, *n* (%)	
In situ only	5 (13.8%)
In situ and invasive	22 (61%)
Invasive only	8 (22.2%)
Histology, *n* (%)	
Ductal	27 (75%)
Lobular	6 (16.7%)
Mucinous	1 (2.78%)

Table 4 (continued) | Characteristics of cancers detected by supplemental MRI (*n*=36)

	Median (IQR) or *n* (%)
Grade, *n* (%)	
1	3 (8.3%)
2	27 (75%)
3	5 (13.9%)
Receptor expression, *n* (%)	
HR^+	33 (91.7%)
$HER2^+/HR^-$	0 (0%)
$HER2^-/HR^-$	2 (5.6%)
Ki-67, *n* (%)	
Low	10 (27.8%)
Intermediate	14 (38.9%)
High	6 (16.7%)
Lymph node, *n* (%)	
None excised	14 (38.9%)
Metastasis	3 (8.3%)
No metastasis	19 (52.8%)

proof-of-principle, the AI tool was trained only on mammography images from Hologic equipment, using images of high quality assessed by highly experienced radiologists will need to be validated for other equipment for the method to be generalized. Additionally, the trial did not explicitly assess the efficacy of AISmartDensity in women with a history of cancer. The small number of cancer cases also hinders the possibility of conducting subgroup analyses related to cancer characteristics and the incomplete results from MRI follow-up of the 24 BI-RADS 3–5 lesions have the potential to bias results towards more conservative estimates. Furthermore, it should be noted that the applicability of this study may be limited to centers or regions with similar breast cancer screening programs and intervals. Additionally, the women who chose to participate in the study may have a higher incidence of cancer compared with the average population, for example, caused by being more aware of the risk of cancer due to having a higher frequency of cancer in their family history[31,32]. However, women with a strong family history of breast cancer or genetic mutation are all excluded, as are women who have already noticed a lump in their breast. It is important to clarify that our study exclusively reports on the prevalence round of supplemental MRI screening, that is, the first time this approach is used. For our study, as for any prevalence screening, one would expect a higher cancer detection rate compared with previous and subsequent screening rounds. We would also expect that the stage distribution of screen-detected cancers is shifter towards lower stages compared with previous mammography-only rounds and will be further shifted towards lower stages in subsequent rounds.

In conclusion, using an approach of analyzing mammograms with AI as a selection method for supplemental MRI is around four times more efficient, in terms of cancer detection, than second-best approaches based on traditional density measures and risk models. Subsequent follow-up research will determine primary outcomes of the effect on prognostically important cancer characteristics.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-024-03093-5.

References

1. Marmot, M. et al. The benefits and harms of breast cancer screening: an independent review. *Br. J. Cancer* **108**, 2205–2240 (2013).
2. Duffy, S. W. et al. Beneficial effect of consecutive screening mammography examinations on mortality from breast cancer: a prospective study. *Radiology* **299**, 541–547 (2021).
3. Tornberg, S. et al. A pooled analysis of interval cancer rates in six European countries. *Eur. J. Cancer Prev.* **19**, 87–93 (2010).
4. Crosier, M. et al. Differences in Ki67 and c-erbB2 expression between screen-detected and true interval breast cancers. *Clin. Cancer Res.* **5**, 2682–2688 (1999).
5. Domingo, L. et al. Tumor phenotype and breast density in distinct categories of interval cancer: results of population-based mammography screening in Spain. *Breast Cancer Res.* **16**, 11 (2014).
6. Houssami, N. & Hunter, K. The epidemiology, radiology and biological characteristics of interval breast cancers in population mammography screening. *NPJ Breast Cancer* **3**, 12 (2017).
7. Gilliland, F. D. et al. Biologic characteristics of interval and screen-detected breast cancers. *J. Natl Cancer Inst.* **92**, 743–749 (2000).
8. Hussein, H. et al. Supplemental breast cancer screening in women with dense breasts and negative mammography: a systematic review and meta-analysis. *Radiology* **306**, e221785 (2023).
9. Kriege, M. et al. Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition. *N. Engl. J. Med.* **351**, 427–437 (2004).
10. Checka, C. M., Chun, J. E., Schnabel, F. R., Lee, J. & Toth, H. The relationship of mammographic density and age: implications for breast cancer screening. *Am. J. Roentgenol.* **198**, W292–W295 (2012).
11. Bakker, M. F. et al. Supplemental MRI screening for women with extremely dense breast tissue. *N. Engl. J. Med.* **381**, 2091–2102 (2019).
12. Boyd, N. F. et al. Mammographic density and the risk and detection of breast cancer. *N. Engl. J. Med.* **356**, 227–236 (2007).
13. Byrne, C. et al. Mammographic features and breast cancer risk: effects with time, age, and menopause status. *J. Natl Cancer Inst.* **87**, 1622–1629 (1995).
14. Ursin, G. et al. Mammographic density and breast cancer in three ethnic groups. *Cancer Epidemiol. Biomark. Prev.* **12**, 332–338 (2003).
15. Boyd, N. F. et al. Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study. *J. Natl Cancer Inst.* **87**, 670–675 (1995).
16. Vacek, P. M. & Geller, B. M. A prospective study of breast cancer risk using routine mammographic breast density measurements. *Cancer Epidemiol. Biomark. Prev.* **13**, 715–722 (2004).
17. Bertrand, K. A. et al. Mammographic density and risk of breast cancer by age and tumor characteristics. *Breast Cancer Res.* **15**, R104 (2013).
18. Geuzinge, H. A. et al. Cost-effectiveness of magnetic resonance imaging screening for women with extremely dense breast tissue. *J. Natl Cancer Inst.* **113**, 1476–1483 (2021).
19. Mann, R. M. et al. Breast cancer screening in women with extremely dense breasts recommendations of the European Society of Breast Imaging (EUSOBI). *Eur. Radiol.* **32**, 4036–4045 (2022).
20. Veenhuizen, S. G. et al. Supplemental breast MRI for women with extremely dense breasts: results of the second screening round of the DENSE trial. *Radiology* **299**, 278–286 (2021).
21. Moore, S. G., Shenoy, P. J., Fanucchi, L., Tumeh, J. W. & Flowers, C. R. Cost-effectiveness of MRI compared to mammography for breast cancer screening in a high risk population. *BMC Health Serv. Res.* **9**, 9 (2009).
22. Mann, R. M., Kuhl, C. K. & Moy, L. Contrast-enhanced MRI for breast cancer screening. *J. Magn. Reson. Imaging* **50**, 377–390 (2019).
23. Liu, Y. et al. Use of an AI score combining cancer signs, masking, and risk to select patients for supplemental breast cancer screening. *Radiology* **311**, e232535 (2024).
24. D'Órsi CJ, S.E. et al. *ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System* (American College of Radiology, 2013).
25. Strand, F. et al. Long-term prognostic implications of risk factors associated with tumor size: a case study of women regularly attending screening. *Breast Cancer Res.* **20**, 31 (2018).
26. Kuhl, C. K. et al. Supplemental breast MR imaging screening of women with average risk of breast cancer. *Radiology* **283**, 361–370 (2017).
27. Chen, S.-Q., Huang, M., Shen, Y.-Y., Liu, C.-L. & Xu, C.-X. Application of abbreviated protocol of magnetic resonance imaging for breast cancer screening in dense breast tissue. *Acad. Radiol.* **24**, 316–320 (2017).
28. Gail, M. H. et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J. Natl Cancer Inst.* **81**, 1879–1886 (1989).
29. Seo, M. et al. Features of undiagnosed breast cancers at screening breast MR imaging and potential utility of computer-aided evaluation. *Korean J. Radio.* **17**, 59–68 (2016).
30. Heywang-Köbrunner, S. H., Hacker, A. & Sedlacek, S. Magnetic resonance imaging: the evolution of breast imaging. *Breast* **22**, S77–S82 (2013).
31. Vallone, F. et al. Factors promoting breast, cervical and colorectal cancer screenings participation: a systematic review. *Psychooncology* **31**, 1435–1447 (2022).
32. Seiffert, K. et al. The effect of family history on screening procedures and prognosis in breast cancer patients - results of a large population-based case-control study. *Breast* **55**, 98–104 (2021).

[1]Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden. [2]Breast Radiology Unit, Karolinska University Hospital, Stockholm, Sweden. [3]School of Computer Science and Technology, Royal Institute of Technology (KTH), Stockholm, Sweden. [4]Science for Life Laboratory, Stockholm, Sweden. [5]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden. [6]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [7]Division of Robotics, Perception, and Learning, Karolinska Institutet, Stockholm, Sweden. ✉e-mail: fredrik.strand@ki.se

Methods

The ScreenTrustMRI trial (ClinicalTrials.gov: NCT04832594) is a randomized trial to evaluate the use of AI to select individuals for supplemental MRI within population-based breast cancer screening at Karolinska University Hospital. Participants in the study were enrolled between 1 April 2021, and 7 April 2023. The current report focuses on predefined secondary outcomes, and an ad hoc cost-comparison analysis, based on a subset of the study population as discussed below.

Study setting

The study was conducted at the breast imaging department of Karolinska University Hospital. Under the Swedish national breast screening program, women between 40 and 74 years old are invited for mammographic screening every 2 years. After the age of 74 years, individuals may sometimes undergo screening on their own request. Screening consists of a two-view full-field digital mammogram of each breast. At Karolinska University Hospital, the screening examination is assessed with nonblinded double-reading (that is, the second reader can view the decision by the first reader). Examinations flagged by either reader are routed to a consensus discussion between two radiologists, who decide whether to recall a woman or not. There are no other methods than mammography for the general population, that is, no tomosynthesis, ultrasound or MRI, and radiologists do not report on mammographic density. All mammograms in the study were obtained on Hologic 3Dimensions mammography equipment.

AISmartDensity—software analyzing mammograms

All screening mammograms at Karolinska University Hospital were analyzed by the AISmartDensity model, which is image analysis software developed collaboratively by investigators at the Karolinska Institute and at the Royal Institute of Technology in Stockholm, Sweden[23]. Briefly, it comprises three types of AI models that utilize previously well-described convolutional neural network approaches for image classification: (1) inherent breast cancer risk—an EfficientNetB3 architecture trained on images of females without current cancer, with the task being to classify females likely to be diagnosed with breast cancer within the next 5 years from those remaining healthy; (2) masking potential—a ResNet34 architecture trained to rank normal mammograms by varying difficulty levels based on radiologists' subjective assessments; and (3) cancer signs—a ResNet34 architecture trained to classify between images with current cancer and those with normal mammograms. AISmartDensity uses an average of standardized scores from the three models. For the cancer signs model, the in-house model was combined with the output of a similar commercial AI (Insight MMG, Lunit Inc.), which also generated conventional breast density measures. The training data for the in-house algorithms were derived mainly from the CSAW dataset, obtained from the Karolinska University Hospital[33]. AISmartDensity is calculated as the age-adjusted sum of the standardized scores from each model. The 'very high' category of AISmartDensity, was calibrated using the retrospective development data to align with the top 8% of the AISmartDensity score, corresponding to a cut-off value of 1.97 (ref. 23).

Eligibility criteria and enrollment

Women who participated in the public screening program who had a standard four-view mammography examination and a 'very high' AISmartDensity score of their screening mammogram were eligible for the trial. Individuals attending screening as part of a high-risk surveillance program (for example, personal history, family history or genetic mutations) undergo supplemental imaging and were therefore excluded from the trial. Additionally, individuals with breast implants, individuals recalled for diagnostic work-up after mammography, and individuals whose images could not be processed by the AI tool were also excluded. Eligible women received an invitation letter and written trial information. Those providing written informed consent were enrolled and randomized either to supplemental breast MRI or as part of an observational control arm, which is not included in the current analysis. Randomization was performed by having a random number generated in Microsoft Excel by a study administrator not involved in the radiological assessments. The randomization procedure did not include stratification or blocks. Participants randomized to the MRI group received an appointment for the examination within 3 months of their initial screening, with a few exceptions. Each participant randomized to MRI only had one screening MRI. For the current report, only individuals who underwent MRI are included in this analysis.

Women randomized to supplementary MRI were sent a questionnaire on breast cancer risk factors and history, as well as potential contraindications to MRI. Individuals who had absolute or relative contraindications against MRI, including those breastfeeding or pregnant, were excluded from having MRI.

MRI protocol

All MRI images were obtained on a GE Signa Premier 3 T MRI scanner. We composed a dedicated MRI protocol for the trial with a total scan time of 12 min. The protocol consists of a localizer, a 2D Axial Fast Spin-Echo T2-weighted Dixon imaging sequence (FA = 111, TE = 102 ms and TR = 5,365 ms, 1.1 × 1.1-mm matrix size and 4-mm slice thickness), a T1-weighted dynamic contrast-enhanced Dixon sequence (FOV = 35 mm, 0.9 × 0.9-mm in-plane and 1.4-mm slice thickness) with view-sharing enabled to achieve desired times. The dynamic contrast-enhanced sequence contained seven timeframes acquired at 0:00, 0:33, 1:04, 1:37, 3:00, 4:30 and 6:00 min after contrast injection. Based on the subtraction images from the fourth time frame, a maximum intensity projection image was produced.

MRI interpretation and diagnostic workflow

MRI images were assessed by two nonblinded breast radiologists and classified according to the ACR MRI BI-RADS classification. For each MRI examination the radiologists assessed fibroglandular proportion (from 'A. almost entirely fat' to 'D. extreme fibroglandular tissue'), corresponding to mammographic density, as well as the background parenchymal enhancement. Any lesions found were classified according to malignancy suspicion of imaging findings (1–5). For examinations with more than one lesion, the BI-RADS score on exam-level was defined by the highest score for any individual lesion. The second reader was instructed to seek consensus by discussing with the first reader for any marked disagreement. Examinations classified as BI-RADS 1 or 2 (negative, benign) were considered healthy and did not undergo further assessment. Examinations with a finding classified as BI-RADS 3, 4 or 5 (probably benign, suspicious, highly suggestive of malignancy) were recalled for a second-look ultrasound. If a lesion could be identified, it was biopsied to obtain a pathology-verified diagnosis. For BI-RADS 3 lesions not visualized on ultrasound, follow-up MRIs at 6, 12 and 24 months were assigned. For BI-RADS 4–5 lesions not visualized on ultrasound, MRI-guided biopsies were arranged. The presence or absence of cancer was confirmed through histopathology. Cases with BI-RADS 4–5 findings were reported to the multidisciplinary team conference and the patient responsibility transferred to the breast surgery department.

Outcomes

Biopsy-verified breast cancer detected by MRI after a negative screening mammography was the primary outcome measure in this report. Breast cancer was defined as any invasive cancer in the breast or ductal cancer in situ. As permitted by the informed consent, we follow each patient for 27 months after diagnosis. For this report, we extracted pathology assessments for the diagnostic biopsy and for the surgical specimen. Also, we recorded the radiologist interpretations of the MRI examinations and subsequent diagnostic work-up. All study participants will have a subsequent follow-up period of 27 months to assess

the total number of cancers with adverse prognostic characteristics, that is, interval cancer, larger than 2 cm invasive cancer, lymph node positive cancer.

Cost-comparison

A high-level cost-comparison analysis was carried out for a range of cancer detection rates from 0 to 120 per 1,000 examinations. An important assumption was that MRI is ten times more expensive per examination compared with mammography. We plotted a curve describing the relative cost per detected cancer compared with mammography. The cancer detection rate, and corresponding relative cost, observed in our trial and that observed in a previous trial using traditional density as the selection mechanism were determined.

Statistical analysis

Cancer detection rate was expressed as number of cancers detected per 1,000 patients receiving MRI, calculated by dividing the number of individuals with pathology-verified breast cancer by the number completing MRI screening, multiplied by 1,000. For PPV we used two metrics. The first metric (PPV1) was based on dividing the number of individuals with pathology-verified cancer by the number with a positive MRI screening result (BI-RADS 3–5). The second metric (PPV3) was based on the denominator of all individuals that underwent biopsy. Stata statistical software v.15.1 was used for all statistical analyses. All statistical tests were two-sided. The level for statistical significance was set at alpha = 0.05. There was no adjustment for multiple testing.

Study oversight

The trial was approved by the ethics review board in Stockholm County and was monitored by an independent Data Safety and Monitoring Committee. The authors assume responsibility for the accuracy and completeness of the data and analyses, and for the adherence of the trial to the study protocol. The study complies with all local and national regulations regarding the use of human study participants and was conducted in accordance to the criteria set by the Declaration of Helsinki. The authors of this manuscript vouch for the accuracy and completeness of the data reported.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

De-identified patient-level prediction scores and diagnostic outcomes will be shared upon request to the corresponding author who will respond within 4 weeks. Other patient-level data, including images, will be shared as far as allowed by applicable regulations including the GDPR and approval by the Karolinska legal department in relation to patient data integrity. An anonymized version of the CSAW dataset involved in training the in-house algorithms is available at https://doi.org/10.5878/45vm-t798.

Code availability

Source code is available at https://github.com/radiology2023/RetrospectiveAISmartDensity.

References

33. Dembrower, K., Lindholm, P. & Strand, F. A multi-million mammography image dataset and population-based screening cohort for the training and evaluation of deep neural networks—the Cohort of Screen-Aged Women (CSAW). *J. Digital Imaging* **33**, 408–413 (2020).

Acknowledgements

We express our gratitude to all patients who participated in the trial, and to all medical professionals who contributed to the acquisition and interpretation of the radiological examinations. We thank MRI radiographers, S. Rasoul Olyaei and C. Srivastava, study administrators E. Annetorp and K. Salo and research engineers F. Cossío Ramirez and L. Cornelio for their contributions to this clinical trial. Funding was provided by Region Stockholm (M. Salim, F.S.), Medtechlabs (Y.L., K.S., H.A., F.S.), the Swedish Breast Cancer Association (F.S.), the Swedish Cancer Society grant no 22-2010-Pj-01-H (F.S.), the Swedish Research Council grant no 2022-01465 (F.S.) and 2020-00692 (M.E.).

Author contributions

M. Salim, D.N. and F.S. performed the investigation. M. Salim, Y.L, M. Sorkhei, D.N., Y.W. and F.S. curated the data. M. Salim visualized the study, performed the formal analysis and wrote the original draft of the manuscript. Y.L., M. Sorkhei, T.F., I.F., M.E., H.A. K.S. and F.S. developed the study methodology, and M. Sorkhei and Y.W. wrote software. T.F., I.F., M.E., H.A., K.S. and F.S. conceptualized the study. Y.W. provided resources. M.E., K.S. and F.S. reviewed and edited the manuscript. F.S. supervised the study.

Funding

Open access funding provided by Karolinska Institute.

Competing interests

F.S. has received speaker fees from Lunit and from Pfizer. F.S. and K.S. are shareholders of ClearScanAI AB. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41591-024-03093-5.

Correspondence and requests for materials should be addressed to Fredrik Strand.

Peer review information *Nature Medicine* thanks Ritse Mann, Savannah Partridge, Jung Hyun Yoon and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Lorenzo Righetto and Saheli Sadanand, in collaboration with the *Nature Medicine* team.

Radiology

ORIGINAL RESEARCH • BREAST IMAGING

Use of an AI Score Combining Cancer Signs, Masking, and Risk to Select Patients for Supplemental Breast Cancer Screening

Yue Liu, PhD • Moein Sorkhei, MSc • Karin Dembrower, PhD • Hossein Azizpour, PhD • Fredrik Strand, PhD • Kevin Smith, PhD

From the Department of Computational Science and Technology (Y.L., M.S., K.S.) and Department of Robotics, Perception and Learning (H.A.), KTH Royal Institute of Technology, Brinellvägen 8, 114 28 Stockholm, Sweden; Science for Life Laboratory, Stockholm, Sweden (Y.L., M.S., K.S.); Department of Physiology and Pharmacology (K.D.) and Department of Pathology and Oncology (F.S.), Karolinska Institute, Stockholm, Sweden; Department of Radiology, Capio Saint Göran Hospital, Stockholm, Sweden (K.D.); and Department of Breast Radiology, Karolinska University Hospital, Stockholm, Sweden (F.S.). Received September 24, 2023; revision requested November 22; final revision received January 26, 2024; accepted February 2. **Address correspondence to** Y.L. (email: *yue3@kth.se*).

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

See also the editorial by Kim and Chang in this issue.

Radiology 2024; 311(1):e232535 • https://doi.org/10.1148/radiol.232535 • Content codes: BR AI

Background: Mammographic density measurements are used to identify patients who should undergo supplemental imaging for breast cancer detection, but artificial intelligence (AI) image analysis may be more effective.

Purpose: To assess whether AISmartDensity—an AI-based score integrating cancer signs, masking, and risk—surpasses measurements of mammographic density in identifying patients for supplemental breast imaging after a negative screening mammogram.

Materials and Methods: This retrospective study included randomly selected individuals who underwent screening mammography at Karolinska University Hospital between January 2008 and December 2015. The models in AISmartDensity were trained and validated using nonoverlapping data. The ability of AISmartDensity to identify future cancer in patients with a negative screening mammogram was evaluated and compared with that of mammographic density models. Sensitivity and positive predictive value (PPV) were calculated for the top 8% of scores, mimicking the proportion of patients in the Breast Imaging Reporting and Data System "extremely dense" category. Model performance was evaluated using area under the receiver operating characteristic curve (AUC) and was compared using the DeLong test.

Results: The study population included 65 325 examinations (median patient age, 53 years [IQR, 47–62 years])—64 870 examinations in healthy patients and 455 examinations in patients with breast cancer diagnosed within 3 years of a negative screening mammogram. The AUC for detecting subsequent cancers was 0.72 and 0.61 ($P < .001$) for AISmartDensity and the best-performing density model (age-adjusted dense area), respectively. For examinations with scores in the top 8%, AISmartDensity identified 152 of 455 (33%) future cancers with a PPV of 2.91%, whereas the best-performing density model (age-adjusted dense area) identified 57 of 455 (13%) future cancers with a PPV of 1.09% ($P < .001$). AISmartDensity identified 32% (41 of 130) and 34% (111 of 325) of interval and next-round screen-detected cancers, whereas the best-performing density model (dense area) identified 16% (21 of 130) and 9% (30 of 325), respectively.

Conclusion: AISmartDensity, integrating cancer signs, masking, and risk, outperformed traditional density models in identifying patients for supplemental imaging after a negative screening mammogram.

Supplemental material is available for this article.

Breast cancer is the most common cancer among women, accounting for approximately 2.26 million reported cases annually (1). To aid early detection, many countries have undertaken population-based mammographic screening, which has decreased breast cancer mortality by up to 30% (2). However, not all cancers are found during mammographic screening due to the limited sensitivity of mammography. In a study assessing eight screening programs, approximately 28% of patients with breast cancer who initially had a negative screening mammogram were diagnosed with breast cancer before the next planned screening 2 years later (3).

Currently, mammographic density measurement is used to select patients for supplemental screening (4–6). While mammographic density is associated with the masking of cancer and increased breast cancer risk (7), using only density likely does not capture all of the predictive information present in mammograms (8).

Artificial intelligence (AI), and recent advances in its use in medical imaging, offers potential for enhancing the accuracy and efficiency of mammogram analysis. Multiple studies and commercial models have focused on the task of detecting cancer signs present in mammograms, aiming to assist or replace the screening radiologist (9). One study demonstrated that AI could be used to select patients for supplemental screening after a negative screening mammogram (10). There have also been efforts to develop AI models to evaluate the likelihood of masking (ie, cancer being concealed by benign structures) (11) or predict the future risk of breast cancer (12,13).

Abbreviations

AI = artificial intelligence, AUC = area under the receiver operating characteristic curve, BI-RADS = Breast Imaging Reporting and Data System, PPV = positive predictive value

Summary

AISmartDensity, an artificial intelligence (AI)–based score combining cancer signs, masking, and risk AI models, outperformed density models in selecting patients who could benefit from supplemental imaging for breast cancer.

Key Results

- In a retrospective study of 65 325 mammographic examinations bootstrapped from 2043 patients, AISmartDensity—combining artificial intelligence models for cancer signs, masking, and risk—identified 33% of future cancers within 3 years, with a positive predictive value (PPV) of 2.91%, outperforming traditional mammographic density with age adjustment, which identified 13%, with a PPV of 1.09% ($P < .001$ for both).
- Among examinations with the highest 8% of scores, AISmartDensity detected 32% and 34% of interval and next-round screen-detected cancers, whereas age-adjusted dense area detected 13% and 12%, respectively.

Moreover, considering cancer signs that may not have been apparent or were misinterpreted at the time of screening is crucial for short-term risk prediction. In a study that retrospectively analyzed mammograms of participants in the Spanish national breast cancer screening program, the authors reported that among 1012 interval cancers, 48.2% were true interval cancers, 23.2% were missed cancers, 17.2% were cancers with minimal signs, and 11.3% were occult cancers (14). This highlights the importance of considering cancer signs as well as other risk factors when selecting patients for follow-up screening.

In this retrospective study, we present an AI-informed score to select patients for supplemental screening, AISmartDensity. AISmartDensity combines AI models that identify cancer signs, masking, and long-term risk. The goal of the study is to assess the effectiveness of the AISmartDensity score in identifying patients likely to have an undetected cancer after a negative mammography screening, and to compare its performance with that of traditional density measures.

Materials and Methods

The regional ethical review board in Stockholm approved this retrospective study and waived the requirement for informed consent.

Study Population

The study population is taken from the Cohort of Screen-Aged Women (15), a population-wide mammographic data set from the Stockholm region of Sweden. The screening program in Sweden invites female patients aged 40–74 years for screening every 18–24 months. We included randomly selected screening examinations among female patients aged 40–74 years who underwent screening mammography between January 2008 and December 2015 at Karolinska University Hospital. All cases of pathologically verified ductal cancer in situ or invasive breast cancer were included, with cancer status updated in 2021 through linkage with regional cancer registers. Stratified bootstrapping was used to align the study population with the Swedish screening distribution with respect to interval and next-round screen-detected cancer. Exclusion criteria included prior cancers, breast implants, and incomplete examinations. Only screening mammograms acquired using Hologic devices were included. To derive the final study population, we also excluded images taken within 2 months of cancer diagnosis, as such cancers are considered screen-detected, and those taken more 3 years before diagnosis, to ensure a follow-up duration that was sufficiently long to allow for the inclusion of all cancers detected in subsequent screening rounds.

Mammographic Density

According to Swedish practice, radiologists assign only a binary read, normal or abnormal, to each screening examination, and Breast Imaging Reporting and Data System (BI-RADS) mammographic density is not assessed. In this study, breast density was calculated with the open-source Laboratory for Individualized Breast Radiodensity Assessment (LIBRA) software, version 1.0.4, from the University of Pennsylvania, which uses mammographic images as input and calculates the density through a learning-based approach (16).

Imaging Procedure

The examinations included bilateral digital mammograms with two standard views, craniocaudal and mediolateral oblique. Images were acquired using Hologic devices during routine mammographic screenings at Karolinska University Hospital; variables such as study date, patient age, laterality, and view position were also recorded.

Model Development

AISmartDensity combines three mammography-based AI models for breast cancer prediction: The first detects cancer signs, the second assesses masking, and the third estimates the risk of developing breast cancer. The cancer signs model (an AI computer-aided diagnosis, or AI-CAD, model) includes a commercial computer-aided design model (Insight MMG; Lunit) and an in-house model trained on various data sets including the Cohort of Screen-Aged Women, the Curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS-DDSM) (17), and the INBreast database (18) (flowchart of training data selection provided in Fig S1). The masking model uses ResNet-34 (19) architecture and was trained on the CSAW-M public data set (11) and mammographic images from Karolinska University Hospital acquired using Hologic devices. The risk model is based on EfficientNet-B3 (20) and was trained on cancer-free mammograms from the Cohort of Screen-Aged Women (flowchart of training data selection provided in Fig S2). All in-house models were trained on data that did not overlap with the study population.

The AISmartDensity score is the mean of the standardized scores from the three models, ranging from 0 to 1. A higher score suggests a higher probability of undetected cancer. The in-house cancer signs and risk models were trained on data from various

Stockholm hospitals and use ensembling of five independent networks, and all models make use of test time augmentation (21). A cross-validation method was applied for these models to improve accuracy.

Evaluation

The performance of AISmartDensity in the study population was compared to that of two mammographic density measures (percent density and dense area) and age-adjusted versions of these measures (from combined age and density logistic regression models).

Four risk categories were established for predictions from each model, mirroring the BI-RADS density standard: very low (below eighth percentile), low (eighth to 49th percentile), high (50th to 92nd percentile), and very high (above 92nd percentile). These categories were sized to reflect the prevalence of the four BI-RADS density categories: fatty, scattered fibroglandular density, heterogeneously dense, and extremely dense.

We evaluated the performance of the traditional density models and AISmartDensity in predicting future cancer in the study population. Further analysis was conducted based on different cancer characteristics, including next-round screen-detected cancer, interval cancer, advanced cancer (which is a composite end point of interval cancer, cancer with invasive components larger than 15 mm, or cancer with lymph node metastasis), and cancers categorized by invasiveness (invasive vs in situ only) and grade (low, medium, or high).

Finally, we explored the setting in which the cancer signs model is clinically deployed as a reader, assessing AISmartDensity in scenarios where the cancer signs model may flag cases, which we termed "forewarning." We specifically analyzed the performance of AISmartDensity in cases not previously identified with the cancer signs model—that is, not in the top 4% of scores from the cancer signs model—based on a prior study showing the effective cancer detection rate of this model to be at this level (22).

Statistical Analysis

Analyses were conducted at the examination level, considering all four image views. To evaluate overall classification performance, the area under the receiver operating characteristic curve (AUC), along with its 95% CI, was calculated. One thousand repetitions of bootstrapping were conducted to calculate CIs. The examinations with the highest 8% of scores in the bootstrapped study population were assigned to the "very high" category (mimicking the proportion of patients assigned to the BI-RADS "extremely dense" category), and these examinations were used to determine sensitivity and positive predictive value (PPV), which are reported with 95% CIs.

We also computed the years of life saved, which is a theoretical measure of the utility of a method for supplemental screening. Years of life saved was calculated as (85 − age at screening) × 0.41, which estimates the maximum theoretical reduction in mortality for individuals with cancer examinations in the top 8% category. It was based on the mean expected age of death for Swedish women (85 years) (23) and the 41% mortality reduction associated with mammographic screening reported in a Swedish screening trial (24). This calculation serves as a basis for comparison between the methods and not as an absolute estimate (25).

The number of examinations positive for next-round screen-detected cancer, interval cancer, cancer with invasive components larger than 15 mm, cancer with lymph node metastasis, cancer with high grade, and cancer of different molecular subtypes was determined using stratified bootstrapping in the study population. We used the Wilcoxon rank sum test to compare age and breast density between patients diagnosed with cancer and those not diagnosed with cancer and the DeLong test for AUC comparisons. For sensitivity and PPV, we applied the χ^2 test. The distributions of the data were assessed using the Shapiro-Wilk test for normality. $P < .05$ was considered to indicate a statistically significant difference. All statistical analyses were conducted by one author (Y.L.) using Python's statistical library SciPy (version 1.11.4).

Further details related to the materials and methods are provided in Appendix S1. The source code is publicly available at *https://github.com/radiology2023/RetrospectiveAISmartDensity*.

Results

Patient Characteristics

The initial cohort for developing the cancer signs and risk models included 57 901 patients. After implementing a threefold cross-validation approach and after exclusions, the training set for the cancer signs model included images from 54 888 patients, and the training set for the risk model included data from 34 306 patients. For the masking model, the training set consisted of 9523 patients. The study population included examinations from 2043 patients held out from the training data, as seen in the flowchart in Figure 1. The study population comprised 258 patients with breast cancer and 1785 healthy individuals (Table 1). Among the 258 patients with cancer, there were 130 interval cancers, while 128 cancers were detected during the subsequent round of screening. Bootstrapping was performed to ensure interval cancer and screen-detected cancer rates of 0.2% and 0.5%, respectively, in the study population, resulting in a total of 65 325 examinations: 64 870 in healthy patients and 455 in patients with cancer.

The median age at mammography was 53 years (IQR, 47–62 years) for healthy patients, which was younger than for patients later diagnosed with breast cancer (59 years [IQR, 49–65 years]; $P < .001$). Healthy individuals had a median percent density of 22.2% (IQR, 14.2%–32.5%) and a median dense area of 32 cm^2 (IQR, 23.8–42.6 cm^2). Patients diagnosed with breast cancer had a similar median percent density of 22.8% (IQR, 14.5%–32.6%; $P = .25$) but a higher median dense area of 34.6 cm^2 (IQR, 25.9–43.2 cm^2; $P = .01$).

Overall Diagnostic Accuracy and Performance

AISmartDensity achieved an AUC of 0.72, a sensitivity of 33% (152 of 455), and a PPV of 2.91% for detecting next-round screen-detected cancers or interval cancers (Table 2). These values were higher than those achieved using traditional density measures. The best among the traditional density measures,

age-adjusted dense area, performed significantly worse, with an AUC of 0.61, a sensitivity of 13% (57 of 455), and a PPV of 1.09% (P < .001 for all comparisons). The corresponding receiver operating characteristic curves are provided in Figure 2. The theoretical years of life saved using AISmartDensity was more than double that of the best-performing traditional density measure (dense area).

Categorization of Examinations with AISmartDensity

The numbers of cancers with different characteristics sorted into the four AISmartDensity categories are shown in Table 3. Of 455 cancers in total, 152 (33%) and 197 (43%) were categorized as very high and high, respectively, based on the previous screening mammogram, while only two cancers, both next-round screen-detected cancers, were categorized as very low. AISmartDensity demonstrated consistent sensitivity for sorting mammograms into the "very high" category for next-round screen-detected cancer, interval cancer, and advanced cancer (sensitivity range, 32% [41 of 130] to 37% [96 of 259]).

AISmartDensity surpassed all traditional density-based predictors in predicting (ie, categorizing as very high) advanced cancers and overall cancer (P < .001 for all comparisons). It was also superior at predicting invasive cancers, high-grade cancers, next-round screen-detected cancers, and interval cancers (P < .05 for all comparisons). Considering invasiveness, AISmartDensity detected more future invasive cancers (139 of 405 [34%]) compared with age-adjusted dense area (48 of 405 [12%]; P < .001), although it found fewer in situ cancers (11 of 42 [26%]) compared with percent density (13 of 42 [31%]; P = .003). Considering cancer grade, AISmartDensity found 19 of 116 (16%) future high-grade cancers; the best-performing traditional density measurement, percent density, was competitive, finding 18 of 116 (16%). For medium- and low-grade cancer, AISmartDensity outperformed the traditional models. AISmartDensity identified 83 of 227 (37%) future medium-grade cancers, while the best-performing traditional density measurement (age-adjusted dense area) detected 36 (16%). AISmartDensity detected 36 of 93 (39%) future low-grade cancers, while age-adjusted dense area detected only six of 93 (6%).

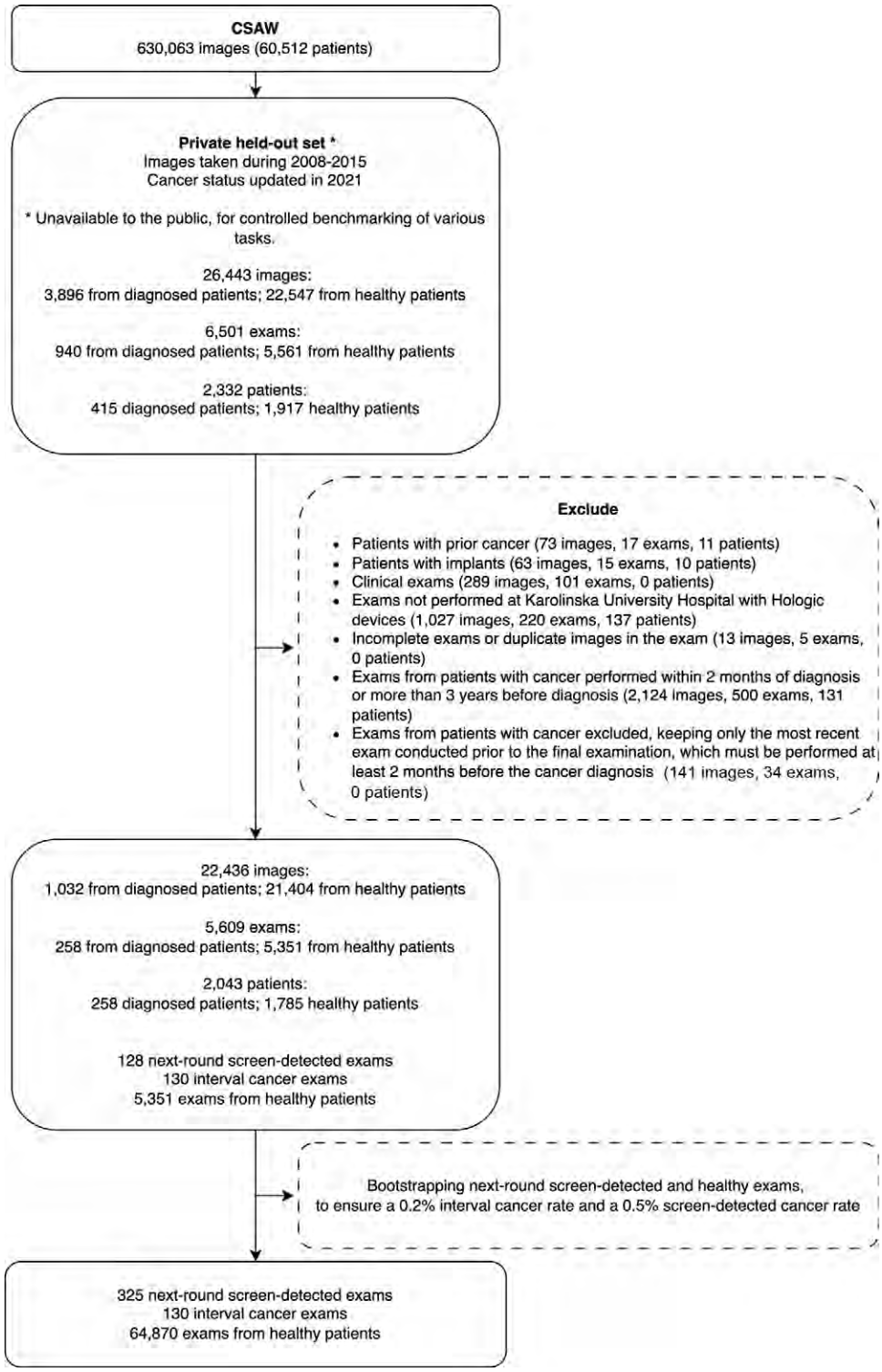

Figure 1: Study population flowchart. CSAW = Cohort of Screen-Aged Women.

Table 1: Characteristics of the Study Population

Characteristic	All Patients	Healthy Patients	Patients Diagnosed with Breast Cancer	*P* Value
No. of examinations	65 325	64 870	455	
Age at mammography (y)*	53.0 (47.0–62.0)	53.0 (47.0–62.0)	59.0 (49.0–65.0)	<.001
Percent density (%)*	22.2 (14.2–32.5)	22.2 (14.2–32.5)	22.8 (14.5–32.6)	.25
Dense area (cm^2)*	32.0 (23.8–42.6)	32.0 (23.8–42.6)	34.6 (25.9–43.2)	.01
Tumor characteristics in patients later diagnosed with cancer				
Detection mode				
Next-round screen-detected cancer			325 (71)	
Interval cancer			130 (29)	
Invasiveness				
In situ cancer only			42 (9)	
Invasive cancer			405 (89)	
Missing information			8 (2)	
Histopathologic size of the invasive cancer				
≤15 mm			226 (50)	
>15 mm			159 (35)	
Missing information			70 (15)	
Metastasis to lymph nodes				
Yes			102 (22)	
No			338 (74)	
Missing information			15 (3)	
Grade				
Low			93 (20)	
Medium			227 (50)	
High			116 (25)	
Missing information			19 (4)	
Histologic type				
Ductal			333 (73)	
Lobular			46 (10)	
Other			31 (7)	
Missing information			45 (10)	
Molecular subtype				
Luminal A			276 (61)	
Luminal B			37 (8)	
Her2-overexpressing			16 (4)	
Basal			15 (3)	
Missing information			111 (24)	

Note.—Categorical data are presented as numbers of examinations, with percentages in parentheses. The study population was bootstrapped from 1785 examinations in healthy patients and 258 examinations in patients diagnosed with breast cancer, to be in line with population rates of interval and next-round screen-detected cancer. Cancer status was ascertained through a 36-month follow-up time, excluding screen-detected cases diagnosed during the first 2 months after screening mammography.

* Continuous data are presented medians, with IQRs in parentheses. The two groups were compared using the Wilcoxon rank sum test (*P* value).

Performance Excluding Examinations Forewarned by the Cancer Signs Model

In Table 4, we repeated the analysis shown in Table 2—in which AISmartDensity was compared against the density measures in terms of AUC, sensitivity, PPV, and years of life saved—but with forewarned examinations (ie, the top 4% from the cancer signs model) excluded from the analysis under the assumption they would be identified in routine screening. There was a noticeable drop in the performance of AISmartDensity, but it still outperformed traditional density measures in terms of AUC (0.67 vs 0.54–0.60; $P < .001$ for all), sensitivity (21% vs 9%–13%; *P* value range, <.001 to .01), and PPV (1.46% vs 0.64%–0.94%; *P* value range, <.001 to .02); theoretical years of life saved was 761 for AISmartDensity versus 576 for the best-performing density measure.

Finally, for a breakdown of the performance of the components of AISmartDensity, refer to Appendix S1 and Table S1. For examples of mammograms from the study population

Table 2: Performance Metrics for Detecting Subsequent Cancers

Predictor	AUC		Sensitivity (%)†		PPV (%)		Years of Life Saved‡
	Value*	*P* Value	Value	*P* Value	Value	*P* Value	
Percent density	0.51 (0.51, 0.51)	<.001	9 (40/455)	<.001	0.77 (40/5226)	<.001	598
Age-adjusted percent density	0.60 (0.60, 0.60)	<.001	9 (40/455)	<.001	0.77 (40/5226)	<.001	307
Dense area	0.53 (0.53, 0.53)	<.001	11 (51/455)	<.001	0.98 (51/5226)	<.001	675
Age-adjusted dense area	0.61 (0.61, 0.61)	<.001	13 (57/455)	<.001	1.09 (57/5226)	<.001	455
AISmartDensity	0.72 (0.72, 0.72)	Ref.	33 (152/455)	Ref.	2.91 (152/5226)	Ref.	1579

Note.—Except where noted, data in parentheses are numbers of examinations. The highest 8% of scores (ie, those categorized as very high) were used to determine the sensitivity, PPV, and years of life saved. *P* values are from statistical tests (DeLong test for AUC and χ^2 test for sensitivity and PPV) comparing the result for each predictor with that of AISmartDensity. AUC = area under the receiver operating characteristic curve, PPV = positive predictive value.

* Data in parentheses are 95% CIs.

† Total cancers bootstrapped from 258 (130 interval cancers and 128 next-round screen-detected cancers) to 455 (130 interval cancers and 325 next-round screen-detected cancers).

‡ Years of life saved estimates the total additional years of life expectancy gained aggregated among patients identified as very high risk by the model.

categorized by the component models (cancer signs, masking, and risk), see Figure S4.

Discussion

In screening programs that offer supplemental imaging after negative screening mammogram, the current method used to select participants is based on traditional mammographic density measures. However, summary measures of density alone may not capture all predictive information available in mammograms (8). Therefore, we developed AISmartDensity, a composite score that integrates artificial intelligence models of cancer signs, masking, and the risk of breast cancer. By selecting the 8% of examinations with the highest AISmartDensity scores among patients with negative mammograms, we could more correctly identify patients who would be diagnosed with cancer within the next 3 years (with a sensitivity of 33%, a positive predictive value [PPV] of 2.91%, and an area under the receiver operating characteristic curve [AUC] of 0.72) than when using age-adjusted dense area (with a sensitivity of 13%, a PPV of 1.09%, and an AUC of 0.61; $P < .001$ for all comparisons). AISmartDensity would potentially save over twice the number of years of life compared with the best-performing density measure (1579 years vs 675 years).

Density is commonly used to simultaneously convey information about the risk of breast cancer and the potential masking of cancer. A higher density could warrant the use of MRI for screening, which has higher sensitivity but is more costly and time-consuming. Age-adjusted density models have been shown to be predictive of breast cancer according to several studies (7,26). In our study, we found an AUC of 0.61 for the age-adjusted dense area model, which is in line with the AUC values of 0.58 and 0.55 found in previous studies (27,28).

We found that AISmartDensity score, which combines the output scores of three AI models, greatly outperformed density-based models in predicting short-term breast cancer diagnosis ($P < .001$). This enhanced performance persisted when forewarned examinations (the top 4% of examinations from the cancer signs model) were excluded, illustrating the effectiveness of AISmartDensity even in a practical scenario where the cancer signs model may be employed in routine screen-reading. In clinical practice, individuals with very high AISmartDensity score may benefit from supplemental screening tests to enhance early detection. Moreover, this risk categorization might also prompt an additional image review by radiologists, particularly when informed of a high AISmartDensity score.

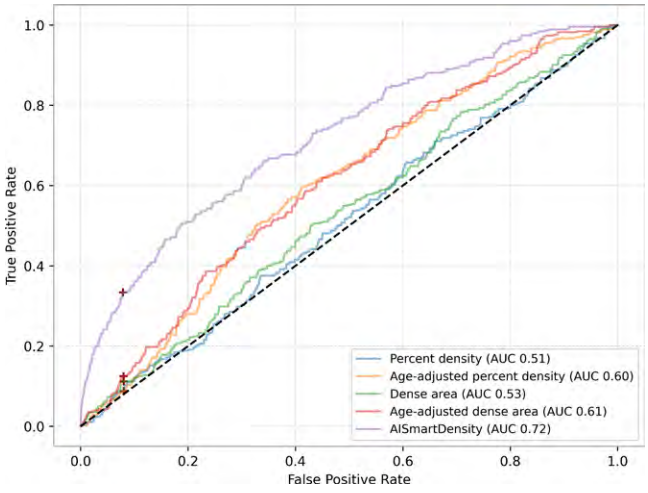

Figure 2: Receiver operating characteristic curves for traditional mammographic density models and an artificial intelligence–based score, AISmartDensity, for detecting subsequent cancers. + = operating point, AUC = area under the receiver operating characteristic curve.

Apart from being the best in selecting which examinations were likely to be followed by cancers of any kind, AISmartDensity outperformed traditional density indicators in selecting which examinations were likely to be followed by interval cancer or advanced cancer ($P < .001$ for all comparisons). Using the AISmartDensity "very low" category to suggest which patients could skip the subsequent screening round without posing a major risk would achieve a very low false-negative rate of 0.44%,

Table 3: Detection of Cancers with Different Characteristics

Predictor	Any Cancer (*n* = 455)	Advanced Cancer (*n* = 259)*	Invasiveness		Grade			Detection Mode	
			In Situ Cancer Only (*n* = 42)	Invasive Cancer (*n* = 405)	Low (*n* = 93)	Medium (*n* = 227)	High (*n* = 116)	Next-Round Screen-Detected (*n* = 325)	Interval Cancer (*n* = 130)
Percent density									
Very low	36 (8)	18 (7)	4 (10)	32 (8)	10 (11)	12 (5)	14 (12)	31 (10)	5 (4)
Low	182 (40)	93 (36)	9 (21)	170 (42)	45 (48)	96 (42)	38 (33)	145 (45)	37 (28)
High	197 (43)	126 (49)	16 (38)	177 (44)	33 (35)	106 (47)	46 (40)	127 (39)	70 (54)
Very high	40 (9)	22 (8)	13 (31)	26 (6)	5 (5)	13 (6)	18 (16)	22 (7)	18 (14)
Age-adjusted percent density									
Very low	15 (3)	7 (3)	0 (0)	15 (4)	2 (2)	7 (3)	5 (4)	8 (2)	7 (5)
Low	143 (31)	72 (28)	13 (31)	128 (32)	29 (31)	52 (23)	56 (48)	104 (32)	39 (30)
High	257 (56)	157 (61)	21 (50)	230 (57)	59 (63)	143 (63)	47 (41)	194 (60)	63 (48)
Very high	40 (9)	23 (9)	8 (19)	32 (8)	3 (3)	25 (11)	8 (7)	19 (6)	21 (16)
Dense area									
Very low	30 (7)	9 (3)	4 (10)	26 (6)	8 (9)	12 (5)	10 (9)	25 (8)	5 (4)
Low	174 (38)	98 (38)	5 (12)	167 (41)	46 (49)	89 (39)	36 (31)	131 (40)	43 (33)
High	200 (44)	123 (47)	21 (50)	173 (43)	35 (38)	102 (45)	53 (46)	139 (43)	61 (47)
Very high	51 (11)	29 (11)	12 (29)	39 (10)	4 (4)	24 (11)	17 (15)	30 (9)	21 (16)
Age-adjusted dense area									
Very low	8 (2)	4 (2)	0 (0)	8 (2)	0 (0)	5 (2)	3 (3)	4 (1)	4 (3)
Low	154 (34)	83 (32)	15 (36)	137 (34)	29 (31)	65 (29)	54 (47)	107 (33)	47 (36)
High	236 (52)	132 (51)	18 (43)	212 (52)	58 (62)	121 (53)	48 (41)	174 (54)	62 (48)
Very high	57 (13)	40 (15)	9 (21)	48 (12)	6 (6)	36 (16)	11 (9)	40 (12)	17 (13)
AISmartDensity									
Very low	2 (0)	0 (0)	0 (0)	2 (0)	0 (0)	0 (0)	2 (2)	2 (1)	0 (0)
Low	104 (23)	49 (19)	4 (10)	99 (24)	22 (24)	57 (25)	24 (21)	79 (24)	25 (19)
High	197 (43)	114 (44)	27 (64)	165 (41)	35 (38)	87 (38)	71 (61)	133 (41)	64 (49)
Very high	152 (33)	96 (37)	11 (26)	139 (34)	36 (39)	83 (37)	19 (16)	111 (34)	41 (32)

Note.—Data are numbers of cases, with percentages in parentheses.

* Advanced cancer was defined as interval cancer, cancer with invasive components larger than 15 mm, or cancer with lymph node metastasis.

again highlighting the value of combining all three component models into one score.

Earlier studies used image-based computational models for short-term breast cancer risk prediction. In 2019, Yala et al (12) and our group (13) reported on early AI models predicting breast cancer risk. A comparative study of five models by Arasu et al (29) showed that the best models had AUCs of 0.71 and 0.69 for interval cancer prediction up to 3 years. Our model predicted cancer between 60 days and 3 years after screening with an AUC of 0.72. In terms of clinical relevance, AISmartDensity achieved a detection rate of 29 cancers per 1000 MRI examinations with a top 8% population selection. Lauritzen et al (30) combined a measure of texture risk and examination score, achieving a detection rate of 36 cancers per 1000 MRI examinations with a 10% population selection in a population with a median follow-up time of 38 months (75th percentile, 50 months). A validation study of the iCAD ProFound AI Risk model revealed a detection rate of 23 cancers per 1000 MRI examinations with a 6.2% population selection and 20 cancers per 1000 MRI examinations with a 10% population selection (31). These metrics from retrospective studies can be compared with the results of the DENSE trial (32), a randomized controlled trial in the Netherlands, where selecting patients with the highest BI-RADS density (8% of the Dutch population) for supplemental MRI screening achieved a detection rate of 16.5 cancers per 1000 MRI examinations. Importantly, potential cancer detection in retrospective studies, including ours, may not manifest as the actual cancer detection rate in a real-world setting. While AISmartDensity shares similarities with the above models, it distinguishes itself in that it uses three models targeting crucial factors for predicting future cancers.

A strength of our study is the streamlined three-component model applied, offering simplicity and effectiveness compared with other complex scoring systems. By individually scoring for cancer signs, masking, and risk, the model allows more precise action based on which component score is most pronounced. For example, if the cancer signs score is prominent, it might be worth having the screening radiologist take another look at the

Table 4: Performance Metrics after Exclusion of Cancers Forewarned by the Cancer Signs Model

Predictor	AUC		Sensitivity (%)†		PPV (%)		Years of Life Saved‡
	Value*	*P* Value	Value	*P* Value	Value	*P* Value	
Percent density	0.54 (0.54, 0.54)	<.001	11 (38/353)	<.001	0.76 (38/5017)	.001	572
Age-adjusted percent density	0.60 (0.60, 0.60)	<.001	9 (32/353)	<.001	0.64 (32/5017)	<.001	244
Dense area	0.55 (0.55, 0.55)	<.001	12 (43/353)	.003	0.86 (43/5017)	.007	576
Age-adjusted dense area	0.59 (0.59, 0.59)	<.001	13 (47/353)	.01	0.94 (47/5017)	.02	371
AISmartDensity	0.67 (0.67, 0.67)	Ref.	21 (73/353)	Ref.	1.46 (73/5017)	Ref.	761

Note.—Except where noted, data in parentheses are numbers of examinations. Forewarned examinations (the top 4% from the cancer signs model) were excluded from the analysis. The highest 8% of scores (ie, those categorized as very high) were used to determine the sensitivity, PPV, and years of life saved. *P* values are from statistical tests (DeLong test for AUC and χ^2 test for sensitivity and PPV) comparing the result for each predictor with that of AISmartDensity. AUC = area under the receiver operating characteristic curve, PPV = positive predictive value.

* Data in parentheses are 95% CIs.

† Total cancers bootstrapped from 258 (130 interval cancers and 128 next-round screen-detected cancers) to 455 (130 interval cancers and 325 next-round screen-detected cancers).

‡ Years of life saved estimates the total additional years of life expectancy gained aggregated among patients identified as very high risk using the model.

mammogram. On the other hand, if the masking and cancer risk signals are highlighted, this suggests supplemental imaging would be more appropriate. Furthermore, our modular approach of building separate models for different aspects of the risk-assessment process offers a level of explainability to the AI, as each module can be understood and analyzed independently. In Figure S4, the differences in how each model selects women at high risk is apparent. Another strength is the use of a population-based screening cohort linked to regional cancer registers with details on cancer characteristics, assuring accurate determination of interval cancers.

Our study had certain limitations. First, because the mammographic data used in both model development and validation were obtained from a specific demographic population, time period, and geographic area and were acquired on imaging equipment from a specific vendor, the robustness of our models would need to be validated before being applied in a different setting. Second, the retrospective nature of the study could introduce selection bias and limit the applicability of the findings to prospective settings. Third, our inclusion of only the last mammogram before diagnosis for individuals with cancer while including all available examinations for individuals without cancer could potentially introduce bias and cause us to underestimate the effect of using AISmartDensity, since the AI tool could possibly have identified the need for supplemental imaging in an even earlier mammogram for the women who were diagnosed with cancer. Fourth, since AISmartDensity considers age in the risk model, it tends to identify as high risk patients who are on average older than the screening population; it may be beneficial to recalibrate the model to select younger patients. Fifth, due to the unavailability of race and ethnicity information, our analysis does not account for potential variations in mammogram interpretation across different racial and ethnic groups. Additionally, selecting the top 8% of scores from the bootstrapped study population presents a limitation, as this approach may not fully capture the variability encountered in a broader, real-world population. Last, we gave equal weight to standardized predictions from each model for every examination. One might consider more sophisticated methods for combining the cancer signs, masking, and risk models.

The model is currently being deployed as part of a randomized clinical trial, acting as a postscreening reader to flag high-risk mammograms for supplemental MRI (ClinicalTrials.gov: NCT04832594). This trial will provide further evidence regarding the efficacy of the model in a clinical setting and offer insights into its real-world performance.

In conclusion, AISmartDensity effectively identified patients who were likely to benefit from supplemental imaging after a negative screening mammogram.

Acknowledgments: This work was supported by funding from Region Stockholm and MedTechLabs. We thank Edward Azavedo, Dimitra Ntoula, and Athanasios Zouzos at Karolinska University Hospital for their expertise in assessing mammograms for the training of the masking model.

Author contributions: Guarantors of integrity of entire study, **Y.L., M.S., F.S., K.S.**; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agree to ensure any questions related to the work are appropriately resolved, all authors; literature research, **Y.L., M.S., K.D., F.S., K.S.**; clinical studies, **Y.L., H.A., F.S.**; experimental studies, **Y.L., M.S., K.S.**; statistical analysis, **Y.L., F.S., K.S.**; and manuscript editing, all authors

Disclosures of conflicts of interest: **Y.L.** No relevant relationships. **M.S.** No relevant relationships. **K.D.** Grants or contracts from Bröstcancerförbundet; payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing, or educational events from AstraZeneca and Lunit; member of Svenska Bröstcancergruppen; and vice president of Svensk Förening för Bröstradiologi. **H.A.** Grants or contracts from the Knut and Alice Wallenberg Foundation, Digital Futures, Horizon Europe, and the Swedish e-Science Research Centre and patent pending for the method described in this article. **F.S.** Funding to institution from Region Stockholm and MedTechLabs; grants or contracts to institution from the Swedish Research Council, Swedish Agency for Innovation Systems, Region Stockholm, Swedish Cancer Society, Swedish Breast Cancer Group, and European Union; speaker fees from Lunit and Pfizer; member (unpaid) of the Scientific Advisory Committee of the International Breast Density Workshop; and board member (unpaid) of the Svensk Förening för Bröstradiologi. **K.S.** Grants or contracts from MedTechLabs and Hälsa, Medicin och Teknik (HMT); one patent

filed with the Swedish patent office; and founded a start-up company in Sweden, ClearScanAI, based on the material presented in this article.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71(3):209–249.
2. Tabár L, Vitak B, Chen THH, et al. Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. Radiology 2011;260(3):658–663.
3. Törnberg S, Kemetli L, Ascunce N, et al. A pooled analysis of interval cancer rates in six European countries. Eur J Cancer Prev 2010;19(2):87–93.
4. Hussein H, Abbas E, Keshavarzi S, et al. Supplemental breast cancer screening in women with dense breasts and negative mammography: a systematic review and meta-analysis. Radiology 2023;306(3):e221785.
5. Melnikow J, Fenton JJ, Whitlock EP, et al. Supplemental screening for breast cancer in women with dense breasts: a systematic review for the U.S. Preventive Services Task Force. Ann Intern Med 2016;164(4):268–278.
6. Sprague BL, Gangnon RE, Burt V, et al. Prevalence of mammographically dense breasts in the United States. J Natl Cancer Inst 2014;106(10):dju255.
7. Boyd NF, Guo H, Martin LJ, et al. Mammographic density and the risk and detection of breast cancer. N Engl J Med 2007;356(3):227–236.
8. Jiang S, Colditz GA. Causal mediation analysis using high-dimensional image mediator bounded in irregular domain with an application to breast cancer. Biometrics 2023;79(4):3728–3738.
9. Yoon JH, Strand F, Baltzer PAT, et al. Standalone AI for breast cancer detection at screening digital mammography and digital breast tomosynthesis: a systematic review and meta-analysis. Radiology 2023;307(5):e222639.
10. Dembrower K, Wåhlin E, Liu Y, et al. Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: a retrospective simulation study. Lancet Digit Health 2020;2(9):e468–e474.
11. Sorkhei M, Liu Y, Azizpour H, et al. CSAW-M: an ordinal classification dataset for benchmarking mammographic masking of cancer. arXiv 2112.01330 [preprint] https://arxiv.org/abs/2112.01330. Posted December 2, 2021. Accessed September 18, 2023.
12. Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R. A deep learning mammography-based model for improved breast cancer risk prediction. Radiology 2019;292(1):60–66.
13. Dembrower K, Liu Y, Azizpour H, et al. Comparison of a deep learning risk score and standard mammographic density score for breast cancer risk prediction. Radiology 2020;294(2):265–272.
14. Blanch J, Sala M, Ibáñez J, et al. Impact of risk factors on different interval cancer subtypes in a population-based breast cancer screening programme. PLoS One 2014;9(10):e110207.
15. Dembrower K, Lindholm P, Strand F. A multi-million mammography image dataset and population-based screening cohort for the training and evaluation of deep neural networks—the Cohort of Screen-Aged Women (CSAW). J Digit Imaging 2020;33(2):408–413.
16. Keller BM, Chen J, Daye D, Conant EF, Kontos D. Preliminary evaluation of the publicly available Laboratory for Breast Radiodensity Assessment (LIBRA) software tool: comparison of fully automated area and volumetric density measures in a case-control study with digital mammography. Breast Cancer Res 2015;17(1):117.
17. Lee RS, Gimenez F, Hoogi A, Miyake KK, Gorovoy M, Rubin DL. A curated mammography data set for use in computer-aided detection and diagnosis research. Sci Data 2017;4(1):170177.
18. Moreira IC, Amaral I, Domingues I, Cardoso A, Cardoso MJ, Cardoso JS. INbreast: toward a full-field digital mammographic database. Acad Radiol 2012;19(2):236–248.
19. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. arXiv 1512.03385 [preprint] https://arxiv.org/abs/1512.03385. Posted December 10, 2015. Accessed September 18, 2023.
20. Tan M, Le QV. EfficientNet: rethinking model scaling for convolutional neural networks. arXiv 1905.11946 [preprint] https://arxiv.org/abs/1905.11946. Posted May 28, 2019. Updated September 11, 2020. Accessed September 18, 2023.
21. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Pereira F, Burges CJ, Bottou L, Weinberger KQ, editors. Advances in Neural Information Processing Systems 25 (NIPS 2012). Red Hook, NY: Curran Associates, 2012.
22. Salim M, Wåhlin E, Dembrower K, et al. External evaluation of 3 commercial artificial intelligence algorithms for independent assessment of screening mammograms. JAMA Oncol 2020;6(10):1581–1588.
23. Life expectancy at birth and age 65 by sex 1970–2022 and projection 2023–2070. Statistics Sweden. https://www.scb.se/en/finding-statistics/statistics-by-subject-area/population/population-projections/population-projections/pong/tables-and-graphs/life-expectancy-at-birth-and-age-65-by-sex-and-projection/. Updated April 18, 2023. Accessed January 25, 2024.
24. Duffy SW, Tabár L, Yen AMF, et al. Mammography screening reduces rates of advanced and fatal breast cancers: results in 549,091 women. Cancer 2020;126(13):2971–2979.
25. Wilmink ABM, Quick CRG, Hubbard CS, Day NE. Effectiveness and cost of screening for abdominal aortic aneurysm: results of a population screening program. J Vasc Surg 2003;38(1):72–77.
26. Reeder JG, Vogel VG. Breast cancer prevention. Cancer Treat Res 2008;141:149–164.
27. Barlow WE, White E, Ballard-Barbash R, et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. J Natl Cancer Inst 2006;98(17):1204–1214.
28. Arefan D, Mohamed AA, Berg WA, Zuley ML, Sumkin JH, Wu S. Deep learning modeling using normal mammograms for predicting breast cancer risk. Med Phys 2020;47(1):110–118.
29. Arasu VA, Habel LA, Achacoso NS, et al. Comparison of mammography AI algorithms with a clinical risk model for 5-year breast cancer risk prediction: an observational study. Radiology 2023;307(5):e222733.
30. Lauritzen AD, Rodríguez-Ruiz A, von Euler-Chelpin MC, et al. An artificial intelligence-based mammography screening protocol for breast cancer: outcome and radiologist workload. Radiology 2022;304(1):41–49.
31. Eriksson M, Román M, Gräwingholt A, et al. European validation of an image-derived AI-based short-term risk model for individualized breast cancer screening—a nested case-control study. Lancet Reg Health Eur 2023;37:100798.
32. Bakker MF, de Lange SV, Pijnappel RM, et al. Supplemental MRI screening for women with extremely dense breast tissue. N Engl J Med 2019;381(22):2091–2102.