From the Department of Oncology and Pathology
Karolinska Institutet, Stockholm, Sweden

DEEP LEARNING ALGORITHMS FOR TUMOR DETECTION IN SCREENING MAMMOGRAPHY

Mattie Salim

Karolinska Institutet

Stockholm 2023

Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2023

ISBN 978-91-8016-921-9
Cover illustration: Created by Özlem Dalyan

Deep Learning Algorithms For Tumor Detection In Screening Mammography
Thesis for Doctoral Degree (Ph.D.)

By

Mattie Salim

The thesis will be defended in public at Karolinska University Hospital Solna Stockholm, Torsten Gordh S2:02 at 9.00 am 2nd of June 2023.

Principal Supervisor:
Fredrik Strand
Karolinska Institutet
Department of Oncology and Pathology

Co-supervisor(s):
Martin Eklund
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Kevin Smith
Royal School of Technology / Sci Life Lab
Division of Computational Science and Technology

Theodorous Foukakis
Karolinska Institutet
Department of Oncology and Pathology

Opponent:
Paola Clauser
Medical University of Vienna
Department of Biomedical Imaging and Image-guided Therapy

Examination Board:
Torkel Brismar
Karolinska Institutet
Department of Clinical Sciences, Intervention and Technology

Einar Heiberg
Lunds Universitet
Department of Clinical Physiology

Johan Ahlgren
RCC Uppsala Örebro

I dedicate this to my dear husband **Alexander,** who has believed in me and supported me throughout this journey. I dedicate this to my daughter **Nellie,** for always loving me and making me smile. And last but not least I dedicate this to my mother **Ikdham,** who was the strongest and most hardworking woman I know.

It is during our darkest moments that we must focus to see the light.

- Aristotle Onassis

Popular science summary of the thesis

Breast cancer is the leading cause of global cancer incidence and the most prevalent type of cancer amongst women. In Sweden, about 9,000 women are diagnosed with breast cancer annually.

Mortality rates started to improve in the 1980s and 1990s when many countries introduced screening detection programs. Most women diagnosed are over the age of 50, but younger women can also be affected by the disease. In most cases there are no identifiable risk factors besides age and gender. However other risk factors connected to breast cancer are breast density, heredity, reproductive history, breast feeding, obesity, hormone therapy and alcohol consumption.

Nearly 65% of all breast cancer is detected through mammography screening. In Sweden, all women between the ages of 40–74 are invited for breast cancer screening using mammography every 18–24 months. This screening process involves taking two images of each breast and asking questions regarding clinical breast symptoms such as a new lump or nipple secretion. All screening examinations are reviewed by two breast radiologists. If the mammogram is flagged due to a suspicious finding or if the woman reports clinical breast symptoms, the examination is reviewed at a consensus discussion. Throughout the consensus discussion, a minimum of two experienced breast radiologists review the images and decide whether the woman is healthy or needs further evaluation.

The screening program faces many challenges, a lack of breast radiologist, a discrepancy in examination assessments, a lower mammographic sensitivity for women with dense breasts and a high number of interval cancers. Interval cancers are breast cancers that are clinically detected between two screening examinations, they are usually larger at the time of detection and are linked to a higher rate of mortality. The implementation of artificial intelligence systems in breast cancer screening would help address these challenges by improving the efficiency of the screening process, detecting cancers at an earlier stage and making the screening process more individualized by e.g. offering supplementary examinations to those women with highest risk of developing breast cancer.

The overall aims of my research have been to explore a large retrospective dataset and evaluate the performance benchmarks for radiologists, to compare the diagnostic performance between different AI CAD systems, to examine the differences and similarities in false assessments between AI CAD and radiologists and to examine how artificial intelligence can be used as a triaging tool to select women with the highest need for complementary MRI screening.

In study I, we examined around 1,000,000 screening assessments from radiologists in Stockholm county. We assessed the performance overall and by different tumor characteristics, the benchmarks showed a wide range of performance differences and the sensitivity varied by tumor characteristics.

In study II, we evaluate the performance of three commercial algorithms and compare with the retrospective assessments of the radiologists in study I. Our conclusion is that the best performing algorithm assessed screening mammograms with a diagnostic performance exceeding that of the radiologists.

In study III, we investigate the disagreements in assessments between the artificial intelligence system and the radiologists, with a focus on breast density and tumor characteristics. Our conclusion is that the artificial intelligence system can have an important complementary role when combined with radiologist especially for women with high breast density.

In study IV, we perform a randomized clinical trial to examine the effect of applying deep learning methods to select women for MRI-based breast cancer screening. The trial is still ongoing but so far the results are promising. The interim results indicate that the cancer detection rate is substantially higher than that reported for density-based selection methods.

The results presented in this thesis demonstrate that artificial intelligence is a promising tool for breast cancer detection in screening mammography.

Abstract

Population-wide mammography screening was fully implemented in Sweden in 1997. The implementation has helped to identify breast cancer at earlier stages and thereby lowered mortality by 30-40%. However, it still has its limitations, many studies have shown a discrepancy between radiologist when assessing mammographic examinations. Additionally, women with very dense breasts have a lower mammographic sensitivity and cancers are easily missed. There is also a shortage on breast radiologists and the workload is increasing due to more women being screened. These challenges could be addressed with the help of artificial intelligence systems. The artificial intelligence system can serve both as an assistant to replace one radiologist in a double-reading setting and as a tool to triage women with a high risk of breast cancer for additional screening using other modalities.

In this thesis we used data from two cohorts: the cohort of screen aged women (CSAW) and the ScreenTrust MRI cohort. The primary objectives were to establish performance benchmarks based on radiologists recorded assessments (study I), compare the diagnostic performance of various AI CAD systems (study II), investigate differences and similarities in false assessments between AI CAD and radiologists (study III), and evaluate the potential of artificial intelligence in triaging women for complementary MRI screening (study IV). The data for studies I-III were obtained from CSAW, while the data for study IV were obtained from the MRI ScreenTrust cohort. CSAW is a collection of data from Stockholm County between the years of 2008 and 2015.

Study I was a retrospective multicenter cohort study that examined radiologist performance benchmarks in screening mammography. Operating performance was assessed in terms of abnormal interpretation rate, false negative rate, sensitivity, and specificity. Measures were determined for each quartile of radiologists classified according to performance, and performance was evaluated overall and by different tumor characteristics. The study included a total of 418,041 women and 1,186,045 digital mammograms, and involved 110 radiologists, of which 24 were defined as high-volume readers. Our analysis revealed significant differences in performance between high-volume readers, as well as a variability in sensitivity based on tumor characteristics. This study was presented during the 2019 annual meeting of the Radiological Society of North America, and was awarded the Trainee research prize that same year.

Study II was a retrospective case-control study that evaluated the performance of three commercial algorithms. We performed an external evaluation of these algorithms and compared the retrospective mammography assessments of radiologists with those of the algorithms. Operating performance was determined in terms of abnormal interpretation rate, false negative rate, sensitivity, specificity and the AUC. The study included 8,805 women, of whom 740 women had cancer, and a random sample of 8,066

healthy controls. There were 25 radiologists involved. For a binary decision, the cut-point was defined by the mean specificity of the original first-reader radiologists (96.6%). Our findings showed that one AI algorithm outperformed the other AI algorithm and the original first-reader radiologists. This study was presented during the 2020 annual meeting of the European Society of Radiology.

Study III was a retrospective case-control study that evaluated the differences and similarities in false assessments between an artificial intelligence system and a human reader in screening mammography. In this study we included 714 screening examinations for women diagnosed with breast cancer and 8,003 randomly selected healthy controls. The abnormality threshold was predefined from study II. We examined how false positive and false negative assessments by AI CAD and the first radiologist, were associated with breast density, age and tumor characteristics. Our findings showed that AI makes fewer false negative assessments than radiologists. Combining AI with a radiologist resulted in the most pronounced decrease in false negative assessments for high-density women and women over the age of 55. This study was presented at the 2021 annual meeting of the Radiological Society of North America.

Study IV is a randomized clinical trial that aims to investigate the effect of applying deep learning methods to select women for MRI-based breast cancer screening. The study examines how effectively AI can identify women who should be offered a complementary MRI screening based on their likelihood of having cancer that is not visible on regular mammography. The results reported in this thesis are preliminary and based on examinations from April 1, 2021 to December 31, 2022. During the indicated time period, 481 MRI examinations have been completed, and 28 cancers have been detected, yielding a cancer detection rate of 58.2 per 1,000 examinations. Although, the trial is still ongoing, the inter-rim results suggest that using AI-based selection for supplemental MRI screening can lead to a higher rate of cancer detection than that reported for density-based selection methods.

In conclusion, we have shown that the use of AI for breast cancer detection can increase precision and efficiency in mammography screening.

List of scientific papers

I. **Mattie Salim**, Karin Dembrower, Martin Eklund, Peter Lindholm, Fredrik Strand
Range of Radiologist Performance in a Population-Based Screening Cohort of 1 Million Digital Mammography Examinations
Radiology, 2020; vol 297

II. **Mattie Salim**, Erik Wåhlin, Karin Dembrower, Edward Azavedo, Theodoros Foukakis, Kevin Smith, Martin Eklund, Fredrik Strand
External Evaluation of 3 Commercial Artificial Intelligence Algorithms for Independent Assessment of Screening Mammograms
JAMA Oncology, 2020; vol 6

III. **Mattie Salim**, Karin Dembrower, Martin Eklund, Kevin Smith, Fredrik Strand
Differences and Similarities in False Assessments by AI CAD and Radiologists in Screening Mammography
Manuscript

IV. **Mattie Salim**, Yue Liu, Moein Sorkhei, Martin Eklund, Kevin Smith, Fredrik Strand
Using Artificial Intelligence Computer Aided Detection to Select Women for Supplemental MRI Examinations in Breast Cancer Screening – The ScreenTrust MRI Study – an Interim Report
Manuscript

Scientific papers not included in the thesis

I. Karin Dembrower, Erik, Wåhlin, Yue Liu, **Mattie Salim**, Kevin Smith, Peter Lindholm, Martin Eklund, Fredrik Strand
Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: a retrospective simulation study
The Lancet Digital Health, 2020, vol 2.9

II. Karin Dembrower, **Mattie Salim**, Martin Eklund, Peter Lindholm, Fredrik Strand
Implications for downstream workload and sensitivity based on calibrating an AI CAD algorithm by standalone-reader or combined-reader sensitivity matching
Journal of medical imaging, 2023

Contents

Abbreviations

AI CAD	Artificial intelligence computer aided diagnosis
AI	Artificial intelligence
AIR	Abnormal interpretation rate
AUC	Area under the receiver operating characteristic curve
BCS	Breast conserving surgery
BCRAT	Breast cancer risk assessment tool
BI-RADS	Breast Imaging Reporting and Data System
BRCA1	Breast cancer gene 1
BRCA2	Breast cancer gene 2
CC	Craniocaudal
CEM	Contrast enhanced mammography
CIS	Carcinoma in situ
CDR	Cancer detection rate
CSAW	Cohort of screen aged women
DCIS	Ductal carcinoma in situ
Dnr	Diarienummer
ER	Estrogen receptor
FDA	U.S. Food and drug administration
FNR	False negative rate
FPR	False positive rate
HER2	Human epidermal growth factor receptor 2
HRT	Hormone replacement therapy
IQR	Interquartile range
MLO	Mediolateral oblique
MRI	Magnetic resonance imaging

PPV	Positive predictive value
PR	Progesterone receptor
RAD	Radiologist
RAD 1	First Reader radiologist
RAD 2	Second Reader radiologist
SD	Standard deviation
SNPs	Single nucleotide polymorphisms
TNM	Tumor node metastasis
95% CI	95% confidence interval

1 Introduction

Breast cancer is the most prevalent type of cancer amongst women. In Sweden, approximately 9,000 women are diagnosed with breast cancer annually (1). The corresponding number for women diagnosed globally is 2.3 million (2).

Most women diagnosed are over the age of 50, but younger women can be affected by the disease (1). In most cases there are no identifiable risk factors besides age and gender. However other risk factors connected to breast cancer are breast density, heredity, reproductive history, breast feeding, obesity, hormone therapy and alcohol consumption (3).

Although the breast cancer incidence has been increasing during the last decades, the mortality rate has decreased. Mortality rates started to improve in the 1980s and 1990s when many countries introduced screening detection programs. The implementation has helped to identify breast cancer at earlier stages and thereby lowered mortality by 30-40% (4, 5, 6, 7, 8). However, it still has its limitations, studies have shown a discrepancy between radiologist when assessing mammographic examinations (9), also women with very dense breasts have a lower mammographic sensitivity and cancers are easily missed (10, 11). Studies have shown that the average sensitivity of mammography is significantly lower than the average sensitivity of MRI (12). However, MRI is more costly and time-consuming than mammography, which means that it is not suitable for population-wide screening. Additional limitations in the mammography screening programs is a shortage of breast radiologists and an increasing workload due to more women being screened. The use of AI is gradually being implemented to address these challenges.

The research results presented in this thesis demonstrate that artificial intelligence is a promising tool for detecting breast cancer in screening mammography. Throughout this thesis, AI will be used as an abbreviation for artificial intelligence, more precisely deep learning.

2 Literature review

2.1 The breast

2.1.1 Breast anatomy

The breast is developed from the same embryological tissue in both males and females. It is located over the pectoral muscles in the upper part of the torso and extends from the second to the sixth rib and from the sternum to the midaxillary line (13). The nipple and areola (which is the pigmented area surrounding the nipple), are located at the center of the breast. The breast consists of glandular, stromal and adipose tissue, with varying proportions in each woman, and is divided into four quadrants: upper, lower, lateral and medial. The mammary glands, which are glandular tissue that lies in the superficial fascia are responsible for milk production and secretion. The mammary glands consists of lobes and each breast contains about 15-20 lobes that are further divided into smaller lobules. In addition, each lobule contains groups of milk-producing cells called alveoli. Milk from the alveoli is carried through small ducts that converge into larger ducts, eventually emptying into the nipple. The stromal tissue is mainly composed of collagen, which helps maintain the breasts shape and internal structure (13). The adipose tissue, is located between the glandular tissue and the skin. The amount of fat tissue varies among individuals and can affect the size and shape of the breast. Each breast also contains a network of blood vessels and lymph vessels. The lymph vessels drain into lymph nodes located in groups under the arm (axillary), above the collarbone (supraclavicular) and in the chest (13). The development of the breast is hormone dependent and is inactive until pregnancy starts and milk production is initiated.

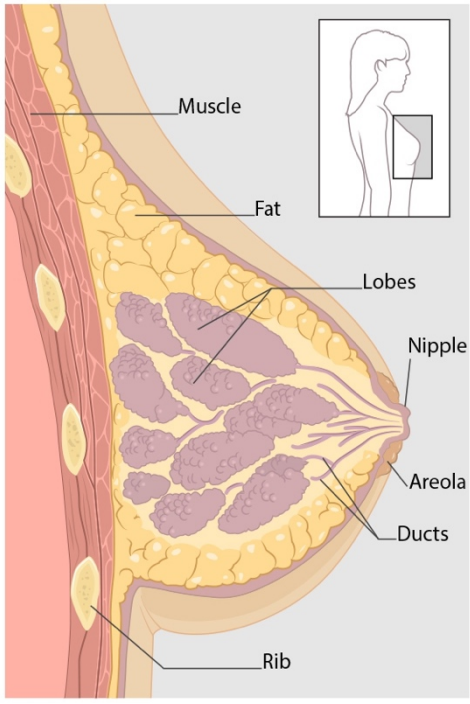


Image 1: Anatomy of the female breast (side view).

(https://www.cdc.gov/cancer/breast/images/breast1_566_838.jpg)

2.1.2 Breast development over time

The female breast is influenced by hormonal changes during the menstrual cycle, pregnancy, breastfeeding and menopause. Before puberty, the breasts are mostly made up of fatty tissue and connective tissue but do not have much functional mammary gland tissue. During puberty, the body experience a surge in hormones, including estrogen and progesterone, which play a key role in the development of the breasts. This causes a proliferation of mammary gland tissue. Proliferation of breast tissue refers to the growth and development of the mammary glands and continues throughout the teenage years and into a woman's early twenties. The development of mammary gland tissue leads to the formation of ducts and lobules.

After breast development is complete, the mammary gland tissue will continue to mature and may undergo further changes during pregnancy and breastfeeding. The proportion of proliferation in the breast tissue varies during puberty, the different stages of the menstrual cycle, pregnancy, lactation and menopause (14, 15).

- During puberty the epithelial cells are immature and undifferentiated with high proliferation.
- During the menstrual cycle, proliferation is low in the early follicular phase and increases as the cycle progresses, peaking in the luteal phase.
- During pregnancy, proliferation increases as the breasts prepare for lactation.
- During lactation, the proportion of proliferation decreases as the breasts are producing milk.
- After pregnancy and lactation, the mammary glands may undergo involution, which is the process of shrinking and returning to their pre-pregnancy state (16).
- During menopause, proliferation decreases as the levels of estrogen and progesterone decrease. At this stage the breast are mainly composed of fatty tissue as the stromal tissue regress (17).

Compared to parous women, nulliparous women have less differentiated breast tissue overall. Studies have shown that lower differentiation in the breast tissue makes it more susceptible to carcinogenesis, which can explain the higher incidence of breast cancer in nulliparous women (16, 18, 19).

2.2 Breast cancer

2.2.1 Epidemiology

As of 2021 breast cancer is the leading form of cancer globally and the incidence is increasing yearly by approximately 0.5% (2, 20, 21). Approximately one in eight women will be diagnosed with breast cancer at some point in their lives and one out of 39 women will die due to the disease (1).

Breast cancer incidence vary all over the world, however it is the most common cancer form in women in both developed and undeveloped regions. Globally 2.3 million women received a breast cancer diagnosis in the year of 2020 (21, 22). In Sweden around 9,000 women are diagnosed with the disease each year (1, 22). In 2020, there were 1,385 deaths for women aged 15 and older due to breast cancer, which accounts for 33 deaths per 100,000 women. The mortality has been reported to be higher for older age groups and less than 5% of women are under the age of 40 when they receive a diagnosis (23). Although breast cancer incidence is increasing, breast cancer related deaths have decreased over time, this could be explained by the introduction of mammography screening programs. Early diagnosis with mammography screening accounts for a 30-40% decrease of breast cancer mortality (4, 5, 6, 7, 8). Additionally, the incidence of breast cancer has had a clear decline for the age group of 50 to 60 years from the year of 2003 and onwards. This is mainly due to the decrease of the use of hormone replacement therapy (HRT) in response to studies being published indicating the connection between HRT and breast cancer (24).

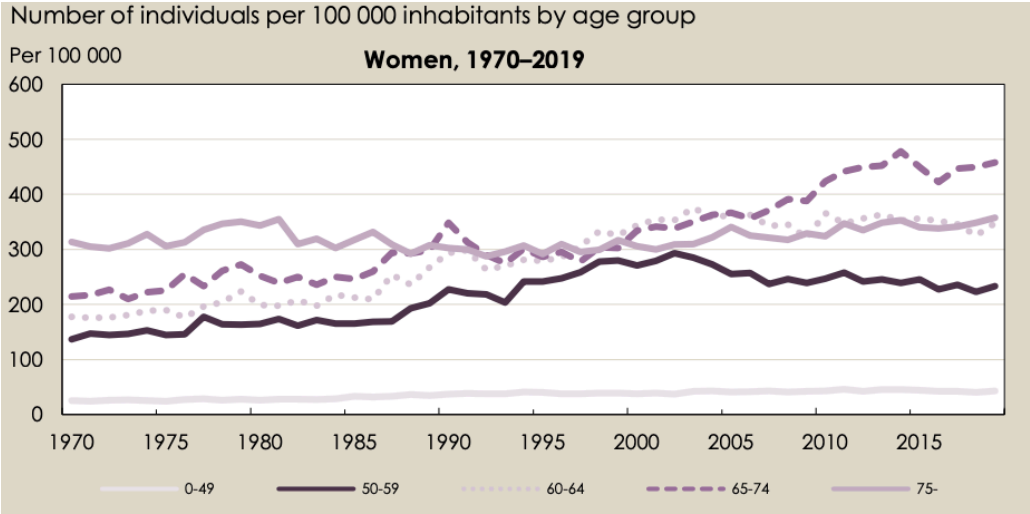


Image 2: Breast cancer incidence over time between age groups during the years 1970-2019.

(https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/artikelkatalog/statistik/2020-12-7133.pdf)

2.2.2 Signs and symptoms

The symptoms of breast cancer vary among women, with the most common sign being the appearance of a lump or mass that feels different from the rest of the breast tissue. Other symptoms of breast cancer may include; skin thickening, change in the size or shape of the breasts, discomfort or pain, dimpling of the skin, a rash, nipple discharge or a change in breast color (25). However, it is important to note that these symptoms can also be caused by benign breast conditions like mastitis and fibroadenomas, which are more common than breast cancer.

2.2.3 Risk factors

Breast cancer has several risk factors, both genetic and non-genetic. The identifiable risk factors are gender, increasing age, family history, obesity, reproductive history, alcohol consumption, tobacco use, history of radiation exposure, mammographic density and postmenopausal hormone therapy (10, 26, 27, 28, 29). However, 50% of breast cancer occur in women who have no known risk factor other than increasing age (>40 years) and gender (female) (27). Studies have shown that the age-related increase in risk slows down after menopause (27, 30, 31, 32).

A woman's risk of getting breast cancer increases if she has a family history of the disease. However, hereditary breast cancer only accounts for 5-10% of all breast cancers (27, 28, 29). Certain inherited gene mutations greatly increase the risk of developing breast cancer by as much as 50-80%, primarily mutations in the breast cancer gene 1 (BRCA1) and the breast cancer gene 2 (BRCA2) (24, 33). Studies indicate that carriers of the BRCA1 gene tend to have more aggressive tumors, such as triple-negative breast cancers, compared to BRCA2 gene carriers who are more likely to have ductal carcinoma in situ. Additionally, carriers of the BRCA1 gene are also more likely to have tumors that are less visible on mammography (34). In Sweden, genetic testing for the BRCA1 and BRCA2 genes is offered to the following patient groups:

- Women with breast cancer <40 years of age,
- Women with triple negative breast cancer,
- Women with breast cancer <50 years of age, and at least one relative with breast cancer,
- Women <60 years of age, with at least two relatives of breast cancer,
- Male breast cancer regardless of age,
- Ovarian cancer patients

After testing, if they are identified as carriers of the mutations they are offered yearly breast imaging that involves magnetic resonance imaging (MRI). Furthermore, two additional groups of inherited gene mutations that are associated with an increased risk of breast cancer have been identified. One of these groups consist of less common genetic mutations, such as CHEK2, PALB2 and ATM, while the other group consists of single nucleotide polymorphisms (SNPs), which are more common types of mutations (35).

Reproductive factors also play an important role in breast cancer risk. Early onset of menstruation and late onset of menopause, as well as a late first pregnancy, are associated with increased risk of breast cancer (27, 30, 36). Additionally, women who have not given birth or have had fewer pregnancies have a higher risk of developing breast cancer. It is estimated that each childbirth lowers the risk by 7%. Breast feeding is also linked to a reduced risk of breast cancer, with approximately 4% lower risk for each year of completed nursing (37).

Furthermore, mammographic density, which will be discussed in the following section, is considered one of the most significant risk factors for breast cancer (26).

Breast Density

Breast density refers to the amount of fibrograndular tissue visible on a mammogram. The fibrograndular tissue appears white on the mammographic images and the fatty tissue appears black. The denser the breast tissue, the harder it is to see abnormalities such as cancer on a mammogram due to the masking effect of the fibrograndular tissue.

Breast density can be classified into four different categories based on the American College of Radiology (ACR) BI-RADS atlas terminology (38):

A. Almost entirely fatty: This type of breast tissue appears mostly black on a mammogram, as fat absorbs very little X-ray radiation. This is the least dense type of breast tissue.
B. Scattered fibrograndular tissue: This type of breast tissue has some density, but it is not uniform and is interspersed with areas of fat.
C. Heterogenously dense: This type of breast tissue has a mix of dense and fatty tissue, making it more difficult to see abnormalities on a mammogram.
D. Extremely dense: This is the densest type of breast tissue and appears white on a mammogram making it very difficult to detect abnormalities.

This classification is the most frequently used classification for mammographic density (39). Research has shown that women with denser breast tissue are at a higher risk for breast cancer. This can be due to the dense breast tissue making it harder to detect cancer but also due to the breast tissue being more biologically active, which could also increase the risk of cancer (26, 40, 41, 42, 43, 44, 45). In 2006 McCormack et al. conducted a meta-analysis of publications on mammographic patterns in relation to breast cancer risk and found that high mammographic density is strongly associated with an increased breast cancer risk (10, 26). Women with higher breast density (>75% dense tissue) have a two to six times greater risk of developing the disease as opposed to women with lower breast density (<5% dense tissue) (10, 26). Breast density is currently not reported on as part of the screening protocol in Sweden.

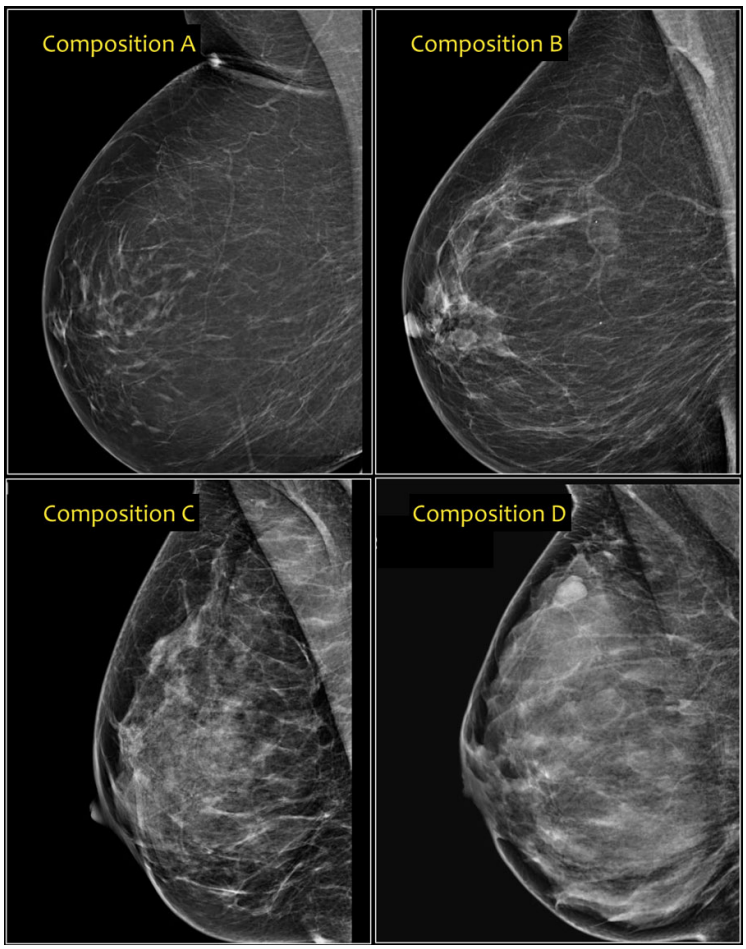


Image 3: Mammographic breast density classification according to ACR BI-RADS atlas.

(https://radiologyassistant.nl/breast/bi-rads/bi-rads-for-mammography-and-ultrasound-2013)

2.2.4 Staging

Tumor stage is used for the clinical assessment of breast cancer to help guide the choice of treatment and determine the prognosis (46). For classification of solid tumors, including breast cancer, the tumor node metastasis (TNM) classification system is used. The TNM system is an international classification system maintained and revised by the American joint committee on cancer and the international union for cancer control (47). The system is the internationally accepted standard not only for breast cancer staging but for staging histologically confirmed carcinomas.

T – Tumor. Used to describe the size of the primary tumor and its' invasion into surrounding tissues. T0 indicates that no evidence of tumor is present, while T1-T4 are used to identify the size and extension of the tumor, with progressive enlargement and invasiveness from T1 to T4.

N – Nodes. Corresponds to the extent of regional lymph node involvement of the tumor. N-values are assessed differently for specific tumors and their regional lymph node drainage. For breast cancer N0 indicates no spread to the lymph nodes. N1 indicates spread to 1-3 axillary lymph nodes. N2 indicates spread to 4-9 axillary lymph nodes and N3 indicates spread to > 9 axillary lymph nodes as well as the infra- and supraclavicular and/or parasternal lymph nodes.

M – Metastasis. Corresponds to the extent of distant metastases of the primary tumor. Metastasis is when the tumor spreads beyond regional lymph nodes. A tumor is classified as M0 if no distant metastasis is present and M1 if there is evidence of distant metastasis (48).

Another staging system that has been shown to correlate strongly with prognosis is the Nottingham histologic grade, also known as the Elston grade. It is a morphological assessment of the tumor, where its gland formation, nuclear image and mitotic activity are combined into a tumor grade of 1-3 (49).

2.2.5 Histopathology and molecular classifications

The histopathology and molecular subtype has become widely important when choosing the appropriate oncological treatment for breast cancer. Breast cancer is divided into non-invasive (carcinoma in situ) and invasive cancer.

Carcinoma in situ (CIS) is responsible for 10.9% of all newly diagnosed breast cancers in Sweden, of which 83% are ductal carcinoma in situ (DCIS) (1). Carcinoma in situ refers to the malignant cells being confined to the ducts or lobules without invasion of the surrounding tissues, this is the initial stage of cancer and generally causes no symptoms. In situ cancers have a low potential for metastasis, however, they may progress over time and invade the surrounding breast tissue and thereby become invasive breast cancer (50, 51). On the contrary, DCIS can also be present and asymptomatic without evolving into invasive cancer for prolonged periods of time. The likelihood of DCIS progressing into a fully invasive cancer is currently uncertain. Untreated cases of DCIS may not always progress due to genotype differences that inhibit progression (52). The classification of DCIS lesions correlate with the clinical course of the disease and is based on the differentiation and growth of the lesion. The lesions are categorized as low-grade, intermediate-grade, or high-grade (53). DCIS is bilateral in approximately 20% of cases (54).

Invasive breast cancer is characterized by the infiltration of cancerous cells into nearby tissues. Invasive breast cancer has the potential to spread to nearby lymph nodes in the axilla (also known as regional metastasis) or to other organs in the body (also known as distant metastasis) (55). The most common type of breast cancers is invasive ductal carcinoma, which is currently classified as invasive carcinoma of no special type according to updated definitions by the world health organization. This group accounts for 70-80% of all breast cancers (56).

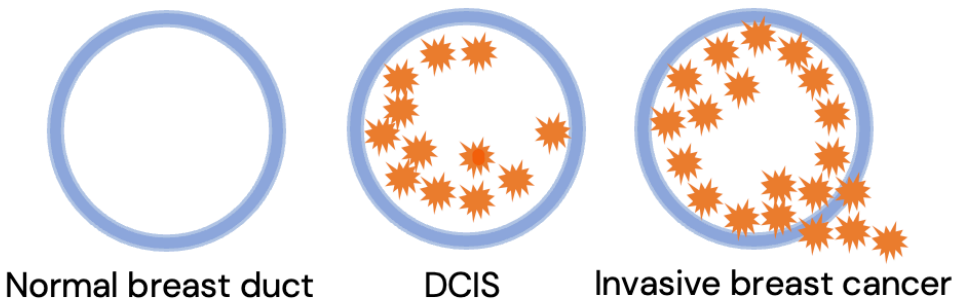


Image 4: Illustration of DCIS and Invasive breast cancer.

Four immunohistochemical biomarkers are being used today in clinical assessment of breast cancer: estrogen receptor (ER), progesterone receptor (PR), HER2 receptor and Ki67.

Approximately 80% of the breast tumors express ER and this expression is strongly correlated with the therapeutic response to endocrine treatment. Breast cancers that are ER-positive are likely to have a more favorable outlook due to their slower rate of proliferation compared to ER-negative cancers (57, 58).

PR expression is mainly of prognostic value. When ER is positive, PR serves as a positive prognostic factor as it is influenced by estrogen. The presence of PR is also linked to the positive response to endocrine and chemotherapy treatment (59, 60).

HER2 is a predictive factor that is linked to more advanced disease, an increased likelihood of relapse and reduced patient survival (61).

Ki-67 is regarded as an unfavorable prognostic indicator. Many studies have demonstrated that a higher proportion of Ki-67 is linked to a reduced survival and a higher chance of tumor recurrence among patients (62, 63). However, in some types of tumors, high Ki-67 levels may indicate a favorable response to chemotherapy.

The St. Gallen classification is a classification system based on the expression of these four immunohistochemical biomarkers and is widely used by clinicians within breast cancer care to help guide and customize treatment. We used the St. Gallen classification as a standard for defining the molecular subtypes in our studies. The classification specifies four different subtypes of breast cancer (64):

- **Luminal A** – This subtype is characterized by positive ER and/or PR, negative HER2 and a low Ki-67 (<14%).
- **Luminal B** – This subtype is characterized by positive ER, negative HER2 and either a high Ki-67 (>14%) or negative PR.
- **HER2-overexpressing** – This subtype is characterized by positive HER2 and negative ER and PR.
- **Basal like** – This subtype is characterized by negative ER, PR, and HER2, and is also known as triple-negative breast cancer.

The subtypes strongly correlate with prognosis, with the luminal A subtype being associated with the most favorable prognosis and the basal like subtype being associated with the least favorable prognosis (65).

2.2.6 Treatment and Prognosis

Breast cancer treatment options are mostly determined by tumor characteristics, patient tolerance and the risk of recurrence. Surgery is the primary choice of treatment for localized breast cancer, either by mastectomy or breast conserving surgery (BCS).

During mastectomy the entire breast tissue is removed. Depending on the patient's medical history and the extent of the cancer, the surgeon may remove the entire breast tissue along with the nipple (total mastectomy), or remove the breast tissue while preserving the nipple (nipple-sparing mastectomy) (66).

The surgical approach most frequently recommended is BCS, which involves removing the tumor along with a surrounding margin of healthy tissue. The amount of tissue removed depends on the size and location of the tumor. Achieving surgical radicality and good functional and cosmetic outcomes are important factors when considering surgery. If the tumor size in relation to breast size makes these outcomes difficult to achieve, preoperative treatment to reduce tumor size should be considered. BCS has the highest rate of success for DCIS and T1-T2 tumors, given that radiation therapy can be administered afterwards. For T1 to T2 breast cancers, BCS followed by radiation therapy has been shown to be just as effective when it comes to survival as a complete mastectomy (67). However, BCS is not recommended for women with a high risk of recurrence (68). It is crucial to achieve tumor-free margins during BCS. In cases of invasive cancers, there should be no presence of tumor on ink, while for DCIS, a tumor-free margin of at least 2 mm is recommended. If this is not obtained, the woman has to undergo re-excision, which happens in about 20% of all BCS cases (69). BCS is always followed by administrating radiotherapy to the entire breast (70), this is also known as adjuvant (post-operative) radiation therapy.

Adjuvant radiation therapy reduces the risk of local recurrence by 50% and increases breast cancer survival rate after both mastectomy and BCS for all breast cancer patients except for patients with low risk tumors and no metastases (70). For these types of tumors no proven benefit of adjuvant radiotherapy has been observed (71). The side effects of radiation therapy can be very troublesome and irreversible and this must be taken into consideration when choosing the appropriate treatment for the patient.

Apart from surgery and radiation, many patients who are diagnosed with high or intermediate risk breast cancer are generally recommended to receive chemotherapy as part of their treatment plan. According to the latest guidelines from the St. Gallen consensus, neoadjuvant chemotherapy, which is a pre-operative systemic treatment, is recommended for all HER2-positive and triple-negative tumors that are 2 cm or larger, even if they are operable (72). For HER2-overexpressing tumors, an anti-HER2 treatment is given in addition to neoadjuvant chemotherapy. An example of this treatment is the monoclonal antibody trastuzumab. Pertuzumab is another monoclonal antibody that is often administered in combination with trastuzumab.

Approximately 80-85% of all diagnosed breast cancer express ER. These receptors bind the female hormone estrogen to the tumor cells and stimulate cell division that leads to tumor growth. Endocrine treatment is a type of cancer therapy that works by preventing ER production or by blocking the action of ER. Tamoxifen, a selective estrogen receptor modulator, works by attaching to ER receptors and blocks the binding of ER to the receptor. Treatment with tamoxifen for five years is the gold standard for pre-menopausal women with ER-positive breast cancer. For post-menopausal women the gold standard is treatment with an aromatase-inhibitor for five years followed by another five years with tamoxifen if there are lymph node metastases present (24).

For locally advanced breast cancers that are inoperable, the recommended treatment is neoadjuvant therapy followed by surgery. However, due to the mammography screening system in Sweden only a few percent of all patients with a primary breast cancer diagnosis have a locally advanced disease. In countries without mammography screening these numbers are significantly higher (73, 74). Patients with locally advanced breast cancer should be offered a treatment consisting of neoadjuvant chemotherapy, surgery and locoregional adjuvant radiation therapy. Women with a ER-positive tumor are in addition recommended a subsequent endocrine treatment. And women with a HER2-overexpressing tumor are offered additional anti-HER2 treatment. There is a lack of recent good-quality clinical trials concerning locally advanced breast cancer, however older studies report approximately a 5-year survival rate of 30-40% following systemic therapy. The corresponding 5-year survival rate for patients without systemic therapy has been reported to be 3.5-15% (75).

The prognosis for breast cancer patients is better than most other cancers, having a 10-year survival rate of around 80% in Sweden (1). So far early detection and treatment has been the largest factor for improvement of breast cancer survival rates (76). Factors that negatively influence prognosis are those connected to advanced disease. One of the main factors connected to metastasis and recurrence is tumor size. Larger tumor size is linked to a higher mortality rate in breast cancer patients (77, 78, 79, 80). Several studies have demonstrated a correlation between tumor size and cancer stage, including lymph node involvement and distant disease (80). Among breast cancer patients, axillary lymph node involvement is the most significant prognostic factor, particularly among women with four or more affected lymph nodes (81). Distant metastasis is another predictor of poor prognosis, with the brain, skeleton, liver and lungs being the most frequently affected sites (80).

2.3 Imaging and diagnostics

2.3.1 Mammography

Mammography is the main breast imaging technique used in breast cancer screening. It is a radiographic examination that plays a central role in early diagnosis of breast cancer. Besides its use in screening it can also be used as a diagnostic tool for women who report breast symptoms outside the time for screening. The image emitted by the mammographic equipment is generated by having an X-ray tube on one side and a detector on the other side of the breast. Mammography uses low X-ray energy (usually around 20 keV) to enhance the contrast between a potential tumor and fatty tissue (82). During the procedure the breast is compressed between two plates in the mammography unit to even out the thickness of the breast tissue and achieve a better image quality.

Mammographic screening has resulted in an earlier detection of tumors which has lowered breast cancer mortality with 20-40% (6, 7, 8, 83). A study by Tabar et al. (8) showed that mortality was lowered by 35% for women that participated regularly in screening. Another, more recent study by Duffy et al. (74) showed that mortality was lowered by 41% within 10 years for women that participated in screening.

The sensitivity of mammography ranges between 58-82% and is in the upper range for women with lower breast density compared to women with higher breast density (26, 40, 84, 85, 86, 87). Diffusely growing cancers such as lobular cancers can be hard to detect on mammography also contributing to a lower mammographic sensitivity (88, 89). Several studies have shown a mammographic specificity that range between 90-95% (87, 90, 91, 92, 93).

The main problems with today's mammography screening is that we fail to identify about 30% of the breast cancers at screening, also known as interval cancers (10, 26, 94, 95). Interval cancers are cancers that go by undetected during screening and are detected later by the woman feeling a lump or a different symptom in the breast such as e.g. pitting of the skin or abnormal nipple discharge. The interval cancer rate is approximately 28% for women regularly attending biennial screening (96). Interval cancers have a higher mortality rate and are more aggressive (97). Besides interval cancers there is also roughly 15% of screen detected cancers that are over 2-cm when detected (98). So despite the decreased mortality due to mammography based breast cancer screening there is still room for improvement.

2.3.2 The mammography screening system in Sweden

Sweden was amongst the first countries in Europe to implement a breast cancer screening program (83, 99). In Sweden the general mammography screening system includes all women 40-74 years of age. The time interval for screening varies between the Swedish counties ranging between 18-24 months. The screening program is organized by the Swedish National Board of Health and Welfare and is offered free of charge to all eligible women. The routine mammography screening acquires a two-view

mammogram that includes a craniocaudal view and a mediolateral oblique view. An illustration of the two views can be seen below. Every screening is reviewed by two independent radiologists – this type of mammogram assessment is referred to as double-reading. If there are any previous mammograms available the radiologist compare them with the current examination. When one or both reviewing radiologists observe a suspicious finding in the mammogram the case gets "flagged" for consensus discussion. The woman gets recalled within seven days for additional procedures if the discussion concludes that there is a suspicion of a pathologic lesion. If the woman reports any breast symptoms such as a lump or nipple discharge at the time of screening she will also be recalled for additional procedures.

Craniocaudal (CC view)

Mediolateraloblique (MLO view)

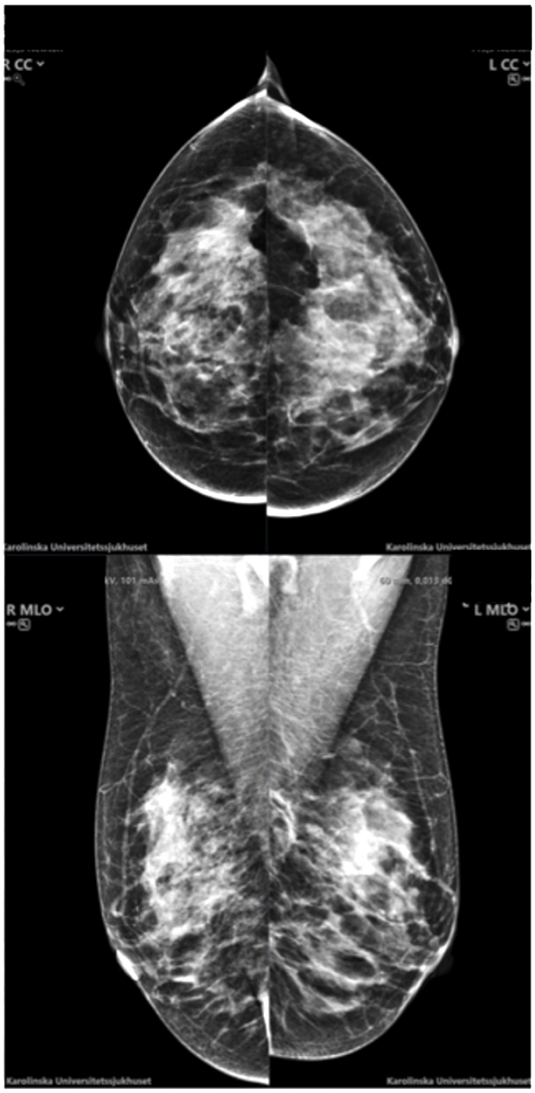


Image 5: Mammographic image, Karolinska university hospital, study participant ScreenTrust MRI study.

In Sweden we use a 5-grade coding system for grading breast lesions in mammographic images, ultrasound examinations and MRI examinations. The European guidelines recommend the five-code classification system that is being used in Sweden where code 1 defines a normal finding and code 5 defines a highly suspicious malignant finding. In North America, the primary system used for mammography is the BI-RADS system, which classifies breast lesions using a seven-code classification system. This system ranges from BI-RADS 1, indicating a normal result, to BI-RADS 5, indicating malignancy. Compared to the Swedish system, the BI-RADS system includes two additional categories: BI-RADS 0 for an incomplete assessment and BI-RADS 6 for a biopsy-verified malignancy. Another difference between the two grading systems is that lesions with code 3 are always biopsied according to the Swedish system while the BI-RADS system allows for the possibility of a six month imaging follow up instead. Category 4 in the BI-RADS system also contains sub-groups (100).

2.3.3 Tumor appearance on mammography

Breast cancer tumors can manifest differently on mammography and can be seen as architectural distortion, microcalcifications, asymmetry and masses that can be spiculated or indistinct.

Architectural distortion is characterized by an alteration in the normal pattern of breast tissue and may be challenging to detect.

Microcalcifications are small white specks that appear on a mammogram, often less than 5 millimeters in size and are composed of calcium deposits that have formed in the breast tissue. When microcalcifications are clustered together in a specific pattern they may indicate a higher likelihood of breast cancer. However, it can be difficult to differentiate them from benign or malignant origin without a biopsy. Clusters of calcifications are typically the most common mammographic features observed in cases of DCIS.

A mass may appear as a denser area in the breast tissue that is distinct from the surrounding tissue. It can have an irregular shape with poorly defined (indistinct) or spiculated edges (spiculated mass) on mammography.

2.3.4 Tomosynthesis

Breast tomosynthesis, also known as 3D mammography has the advantage of reducing the effect of overlapping tissue compared to regular mammography. Hence, tomosynthesis is recommended and mainly used for further evaluation of findings in women that have been recalled from screening to confirm or exclude the presence of a tumor (101). For the evaluation of microcalcifications, tomosynthesis has proven to be just as good as digital mammography. There is sparse literature regarding the use of tomosynthesis in clinical work-up (102), however studies have shown that compared to digital mammography breast tomosynthesis has a higher sensitivity but a lower specificity for tumor detection (103, 104, 105, 106). A recent study by Conant et al. (107) compiled data from five healthcare systems in the United States that used tomosynthesis for breast cancer screening. This study is the largest known study on 3D

mammography. They found that tomosynthesis had a higher cancer detection rate (CDR) and a lower recall rate compared to 2D mammography. Tomosynthesis had a CDR of 5.3 cancers per 1,000 women compared to 4.5 cancers per 1,000 women for 2D mammography.

2.3.5 Contrast Enhanced Mammography

Contrast enhanced mammography (CEM) has emerged as a promising imaging technique for improving the accuracy of breast cancer screening and diagnosis. This technique involves the use of an iodine-based contrast agent injected intravenously prior to mammography. The contrast enhances the visibility of breast lesions and abnormalities, making them easier to detect. This can be particularly useful in mammograms that are difficult to interpret, for example due to dense breast tissue. Several studies have demonstrated the potential benefits of CEM compared to traditional mammography, particularly in cases of dense breast where conventional imaging may have a limited sensitivity. In a systematic review and meta-analysis of 60 studies, the pooled sensitivity was found to be higher than that of traditional mammography (108). Another study that compared the diagnostic performance of CEM and ultrasound with traditional mammography in women with dense breasts and an increased risk of breast cancer found CEM to be more sensitive than traditional mammography but with a reduced specificity (109). Despite many studies showing promising results when compared to traditional mammography (110), CEM is not yet widely adopted in clinical practice. However, when comparing CEM to magnetic resonance imaging (MRI) studies show that MRI is superior in terms of diagnostic performance. A systematic review and meta-analysis published in Radiology aimed to compare the diagnostic performance of CEM to that of contrast enhanced MRI (111). Pötsch et al. examined the diagnostic accuracy of both methods in patients with known abnormalities but no histologic confirmation of cancer. The study results showed that the sensitivity of CEM was significantly lower than that of breast MRI. Although, CEM has certain potential advantages compared to MRI such as; shorter examination time, lower costs and often being more accessible.

2.3.6 Ultrasound

Another imaging technique greatly used within breast cancer diagnosis is breast ultrasound. Breast ultrasound is mainly used as a complement to mammography to examine the breast and axilla, particularly in cases where the woman is recalled from screening. It has also shown promise in women with dense breasts (112, 113, 114) and it can be used as a first hand method for women under the age of 30, pregnant women and lactating women since the technique does not use harmful X-rays. Ultrasound is also used as a guidance tool for breast biopsies. Previously, ultrasound was mostly used to differentiate between cystic and solid tumors but lately the technique has evolved to even be able to detect tumors that are not visible on a mammogram. Studies indicate that the rates of ultrasound detected only cancers range between 2.7 to 4.6 cancers per 1,000 women screened (115, 116, 117, 118, 119, 120, 121, 122). Using ultrasound as a breast

screening method is appealing, since it does not use harmful radiation but the downside is that it is time consuming and operator dependent and also has a lot of false-positive findings (114).

2.3.7 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is used in preoperative assessment of ill-defined tumors to determine the extent of the tumor and as part of screening for gene mutation carriers and women with a family history of breast cancer (123). The examination is performed with the woman lying on her stomach and entering the MRI machine.

Image 6: MRI machine at the mammography department at Karolinska University Hospital

Intravenous contrast media is most often administered during MRI examinations to evaluate the uptake and dynamics of malignant tumors. The intravenous contrast media being used during MRI examinations is Gadolinium based. MRI has been shown to be more sensitive when it comes to breast cancer diagnosis (124). However, it may be perceived as uncomfortable by patients due to a longer and louder examination compared with regular mammography. Additionally, the use of MRI is contradicted for patients with pacemakers and other non-MRI compatible metal in the body, and it might not be suitable for patients with severe claustrophobia. In image 7 you can see an MRI examination with a contrast enhanced lesion in the left breast.

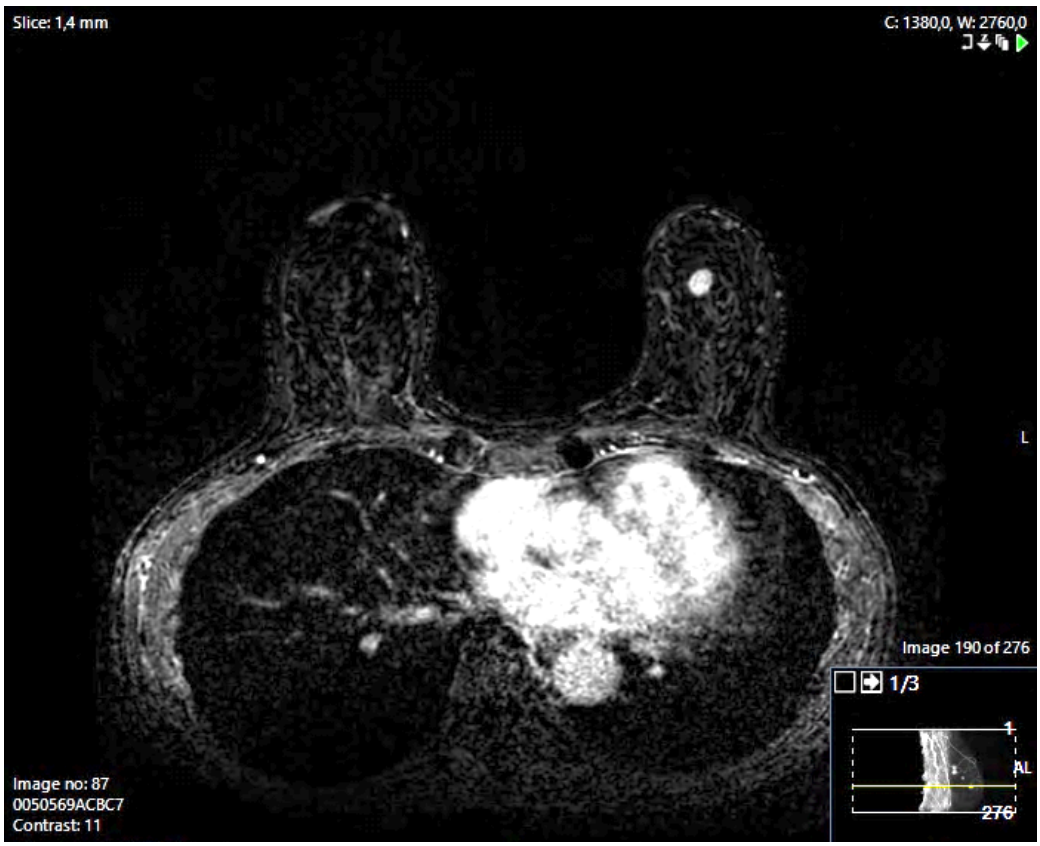


Image 7: An MRI examination image from the ScreenTrust MRI study; consisting of both breasts with a contrast enhanced BI-RADS 5 lesion in the left breast.

The sensitivity for MRI ranges between 80-100%, and the specificity ranges between 83-98% (124, 125, 126, 127, 128, 129, 130). Studies have shown that the number of interval cancers were up to 80% lower when undergoing an MRI examination compared to digital mammography (131, 132). The DENSE trial, a randomized clinical trial by Bakker et al. used breast density as a selection method for complementary MRI examinations in screening. The DENSE trial demonstrated that compared to mammography an MRI examination lowered the number of interval cancers by 80% (132). A recent study by Hussein et al. (133) aimed at evaluating the role of different supplementary screening tests in women with dense breast tissue and average or intermediate risk of breast cancer, who had a negative screening mammogram. The meta-analysis was based on 22 studies including 261,233 patients screened for breast cancer to determine the most effective screening method for women with dense breasts. Breast MRI was the superior screening method, detecting even the smallest cancers with high precision.

Although MRI is the most accurate method for breast cancer diagnosis, traditional mammography is still the only screening modality in wide use today. The downside to MRI is the examination costs, which runs around 5-10 times higher than a mammography examination (134). An additional downside is the long examination time compared to mammography. To address this, new abbreviated MRI protocols have become highly relevant in the development and research of MRI technique for breast cancer screening (135). These protocols reduce radiologist workload and scanner time and do not affect screening accuracy (136, 137). According to a meta-analysis by Baxter et al. (137),

abbreviated MRI did not show a significant decrease in performance in terms of sensitivity or specificity compared to full diagnostic MRI.

To grade MRI examinations, the ACR has established the BI-RADS classification system. We used this classification for grading of all MRI examinations in study IV. The BI-RADS system for MRI classifies findings into seven categories as follows:

- BI-RADS 0 – Incomplete results that require additional imaging evaluation
- BI-RADS 1 – Negative results with no abnormalities found
- BI-RADS 2 – Benign findings such as non-enhancing fibroadenomas and cysts.
- BI-RADS 3 – Propably benign findnings that require follow up within 6 months
- BI-RADS 4 – Suspicious findings that do not have a classic appearance of malignancy but justify a recommendation for biopsy
- BI-RADS 5 – Highly suggestive findings of malignancy that require tissue diagnostic tests
- BI-RADS 6 – Biopsy proven malignancy, which requires MRI for cancer staging or evaluating neoadjuvant therapy

2.3.8 Triple diagnostics

Needle biopsy is done as a complement to breast imaging when there is a suspicion of cancer, this is also known as the triple test, which is the use of three diagnostic modalities for the diagnosis of breast cancer - palpation of the breast, radiologic imaging and pathology. This approach is often referred to as the gold standard when examining all breast abnormalities (138). Using all three modalities gives a high diagnostic safety and improves the likelihood of diagnosis (139, 140). Biopsies are guided by ultrasound, MRI or are vacuum assisted. Core needle biopsies are mainly used when there is a suspicion of cancer while fine needle biopsies are used when there is a suspicious pathological lymph node. The use of vacuum assisted biopsy is becoming more popular as it gives bigger biopsies and increases the chance of getting representative tissue during the procedure. MRI biopsies are most often used when the lesion is not visible or cannot be localized during ultrasound.

2.4 Artificial intelligence in breast imaging

2.4.1 AI, machine learning and deep learning

Artificial intelligence (AI) was developed in the 1950's and has achieved state-of-the-art results in many areas, including healthcare. AI is a computer program that consists of algorithms that can perform complicated tasks that usually require human intelligence without any human intervention.

Machine learning is a subfield of artificial intelligence. Machine learning is the method of teaching computers to learn and make decisions on their own, without being explicitly programmed to perform a specific task. It involves feeding large amounts of data to a computer program, which uses that data to learn how to perform a particular task. Within the machine learning algorithm we have artificial neural networks that are inspired by the structure and function of the human brain. Artificial neural networks are made up of layers of interconnected neurons which process and transmit information.

Deep learning is a subfield of machine learning and is composed of multiple layers of artificial neural networks. Deep learning uses multiple layers of processing to extract higher level features from raw data (141, 142). By combining enough of these layers, deep learning models can learn to recognize and understand complex patterns in the data. Deep learning has surpassed other machine learning methods in various tasks such as natural language understanding, particularly in tasks such as topic classification, question answering and language translation. Additionally, deep learning has advanced in image (143, 144) and speech recognition (142, 145).

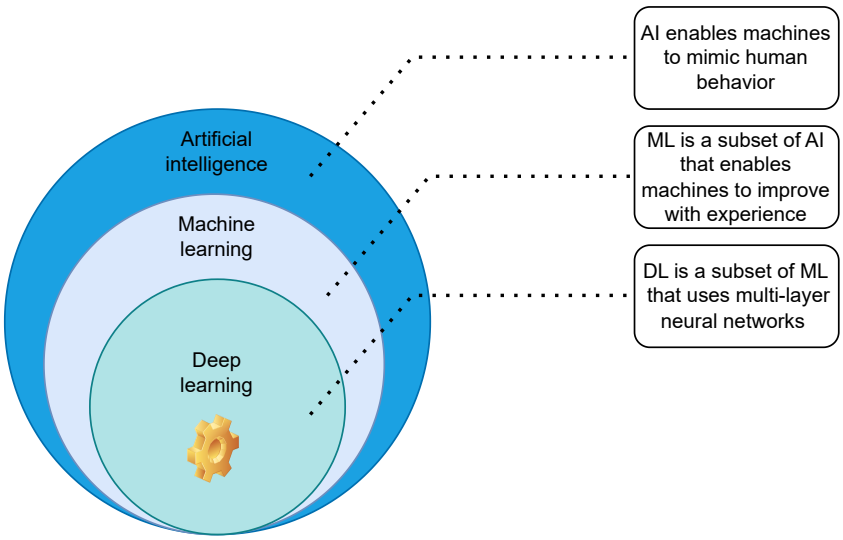


Image 8: Illustration of AI , machine learning and deep learning

Over the past decades there has been a great optimism in developing and implementing artificial intelligence computer aided diagnosis (AI CAD) systems for screening mammography to help aid radiologists in detecting suspicious findings in the images and reduce the broad variation in human performance (146, 147). CAD is based on traditional machine learning techniques and was once widely used within mammography screening. In contrast to modern AI that uses deep learning and is trained on large amounts of data, traditional CAD uses more limited techniques and can only be trained on small amounts of data. Most commercial CAD systems have been used as "assistants" until recently (148). Traditional CAD has had problems with a large number of false-positive findings per mammogram (149). Early evaluation of using CAD as an assistant showed an increased sensitivity but later studies concluded that using CAD did not improve the diagnostic accuracy (148, 150, 151, 152, 153).

2.4.2 Implementation of AI in breast cancer screening

Breast cancer screening programs are widely important for early detection but they phase several challenges. Apart from the wide variation in human performance, the overall costs of breast cancer screening programs are extensive. The cost of breast cancer screening programs can vary depending on the specific program and the country of implementation. For instance, in 2010, the estimated cost for mammography screening in the US was 7.8 billion USD (154). This poses a challenge in low-income countries. The utilization of AI can help enhance the screening performance and reduce the overall costs of mammography screening, thereby mitigating these challenges.

AI can be implemented in various ways, such as an assistant directing the attention of the radiologist to suspicious areas in the image or as an independent reader making an assessment of the mammogram without further human intervention. AI can also be used as a risk predictor to estimate the risk of detecting cancer before the next screening or as a triaging tool triaging cases by complexity.

Recently an increasing number of studies have shown positive results when using artificial intelligence for mammographic tumor detection and also for prediction of future breast cancer (147, 155, 156, 157). A multi-reader, multi-case cross-country study where they compared the stand-alone performance of AI compared to that of 101 radiologists in detecting breast cancer in mammography showed that the performance of the AI system was statistically noninferior to that of the average of the 101 radiologists (147). Another study on AI that shows potential for breast cancer screening, used a large representative data set from the UK and a large enriched data set from the USA. Usage of the AI system showed a reduction in false positives and false negatives. In the same study they ran a simulation in which they used AI in the double-reading process that is used in the UK and found that the AI system not only reduced the workload of the second reader by 88% but also managed to maintain a non-inferior performance to the radiologists (158). Additionally, Leibig et al. (159) conducted a study demonstrating the potential of AI through a decision-referral approach. The study simulated a safety-net warning system that combined traditional triage and cancer detection. Results showed

that this approach was superior to both individual radiologists and a stand-alone AI system in terms of sensitivity and specificity. Conversely, a meta-analysis by Freeman et al. concluded that the use of AI for image analysis in breast cancer screening programs is far from having the quality needed for implementation into clinical practice (160). A total of twelve studies were included and their conclusion was that prospective studies on large screening populations are required before considering an integration of artificial intelligence into clinical practice. This conclusion is on the basis of the studies being of poor methodological quality and that no prospective studies on AI in a mammography screening setting was found.

2.4.3 AI, breast cancer risk and density assessment

There are several tools that can be used to assess an individual's risk of developing breast cancer. These tools take into account different risk factors such as age, family history etc. and use this information to calculate an individual's risk of breast cancer (161). Some examples of the tools being used for breast cancer risk assessment include:

- Gail model: This tool estimates the risk of developing breast cancer over a 5-year period and a lifetime based on factors such as age, family history and personal history of breast biopsies.
- Breast cancer risk assessment tool (BCRAT): This tool estimates the risk of breast cancer over a 5-year period and a lifetime based on riskfactors such as age, family history etc.
- Tyrer-Cuzick model: This tool estimates the risk of breast cancer over a 10-year period and a lifetime based on riskfactors such as age, family history, genetic mutations etc.

However, only the latest version of the Tyrer-Cuzick model takes density into account (44). Density can also be assessed using automated systems. The output of these automated systems is typically a quantitative measure of the breast density.

Recently there have been numerous studies conducted evaluating the risk of breast cancer using AI, and the findings have been promising. The results indicate that AI-based risk assessment tools can provide a more accurate estimate of an individual's risk of breast cancer (162, 163, 164, 165, 166, 167, 168). The algorithms can analyze data based on the above-mentioned factors and in addition identify patterns and features in the images that are associated with an increased breast cancer risk (169). A study by Dembrower et al. evaluated and compared a deep learning risk score with the standardized mammographic density score for breast cancer risk prediction and concluded that the deep neural network could predict which women are at risk for future breast cancer more accurately, with a lower false-negative rate for more aggressive cancers compared to the density-based models. Yala et al. (167) developed a deep learning model that identifies imaging biomarkers on mammograms to predict the risk of developing breast cancer, the model was tested on several diverse data sets from different countries and maintained a high accuracy across all datasets. Another study by Eriksson et al. (168) developed a clinical model that evaluates the short-term

risk by using mammographic density in combination with information regarding several breast cancer risk factors and CAD evaluation of the images. The results showed a high AUC for the full model enabling early identification of women with high risk of breast cancer. Overall, the use of AI in breast cancer risk assessment has the potential to improve the accuracy and efficiency of the risk assessment process.

Additionally, some studies have evaluated the use of AI for breast density assessment. A study by Lehman et al. (170) developed a deep learning algorithm for automatic assessment of breast density. The model achieved high accuracy in classifying breast density with a high sensitivity and specificity. Another study by Magni et al. (171) also developed a deep learning algorithm for breast density classification. Their results revealed that the AI algorithm performed comparably to the radiologists in determining breast density.

3 Research aims

The primary objective of this thesis was to examine the operational performance of AI systems and investigate the potential integration of AI into the mammography screening system.

The specific aims for each study are listed below:

3.1 Study I – Radiologist performance in Breast cancer screening

Our aim was to establish performance benchmarks based on radiologists assessments in the source cohort, facilitating a comparison between the performance of AI CAD systems and these benchmarks.

We wanted to get a better understanding of the performance of the radiologists in our dataset and how the performance was influenced by tumor characteristics.

3.2 Study II – Comparative study of various AI CAD systems as independent readers

Our aim was to determine and compare the performance of three different commercial AI CAD systems as they are applied to make screening decisions without human interaction.

Our hypothesis was that one of the algorithms would exhibit superior performance compared to the other algorithms.

3.3 Study III – Exploring differences between human and AI CAD systems in screening mammography

Our aim was to examine the disagreements in assessments between the AI CAD system and the radiologists.

We hypothesized that variances in evaluations could provide improved insights into the potential outcomes of implementing AI in breast cancer screening.

3.4 Study IV – Examining the effect of applying deep learning methods to select women for MRI based breast cancer screening

Our aim was to perform a randomized clinical trial to examine how effectively AI can identify women that should be offered a complementary MRI screening based on their likelihood of having cancer that is not visible on mammography. One commercial and three in-house-developed AI algorithms were combined to calculate the AISmartDensity score.

Our hypothesis was that by utilizing AI-based selection for supplemental screening rather than density-based selection, we would identify a greater number of undetected cancers.

4 Materials and methods

4.1 Study population

Studies I,II and III were retrospective studies derived from the CSAW dataset (172, 173). CSAW is a dataset containing all women invited for mammographic screening within the Stockholm County area between 2008 and 2015. Prospectively recorded information on radiological assessments and clinical cancer data were extracted from the Regional Cancer Center Stockholm-Gotland. Study I included 418 041 women. Study II and study III included a case-control subset of data within CSAW, based on a random selection of healthy women and all women diagnosed with breast cancer within the data set.

Study IV is a prospective study that includes all women undergoing screening at Karolinska University Hospital from April 1st 2021 until April 7th 2023.

An overview of the study populations and methods for studies I-IV is presented in Table 1.

Table 1.

	Study I	**Study II**	**Study III**	**Study IV**
Study design	Retrospective multicenter cohort study	Retrospective case-control study	Retrospective case-control study	Prospective case-control study
Sample	CSAW dataset	Subset of CSAW dataset	Subset of CSAW dataset	Subset of ScreenTrust MRI dataset
Inclusion period	1/1/2008-31/9/2015	1/1/2008-31/12/2015	1/1/2008-31/12/2015	1/4/2021-31/12/2022
Total women included	418,014	8,805	8,743	481*

*) Women that performed MRI.

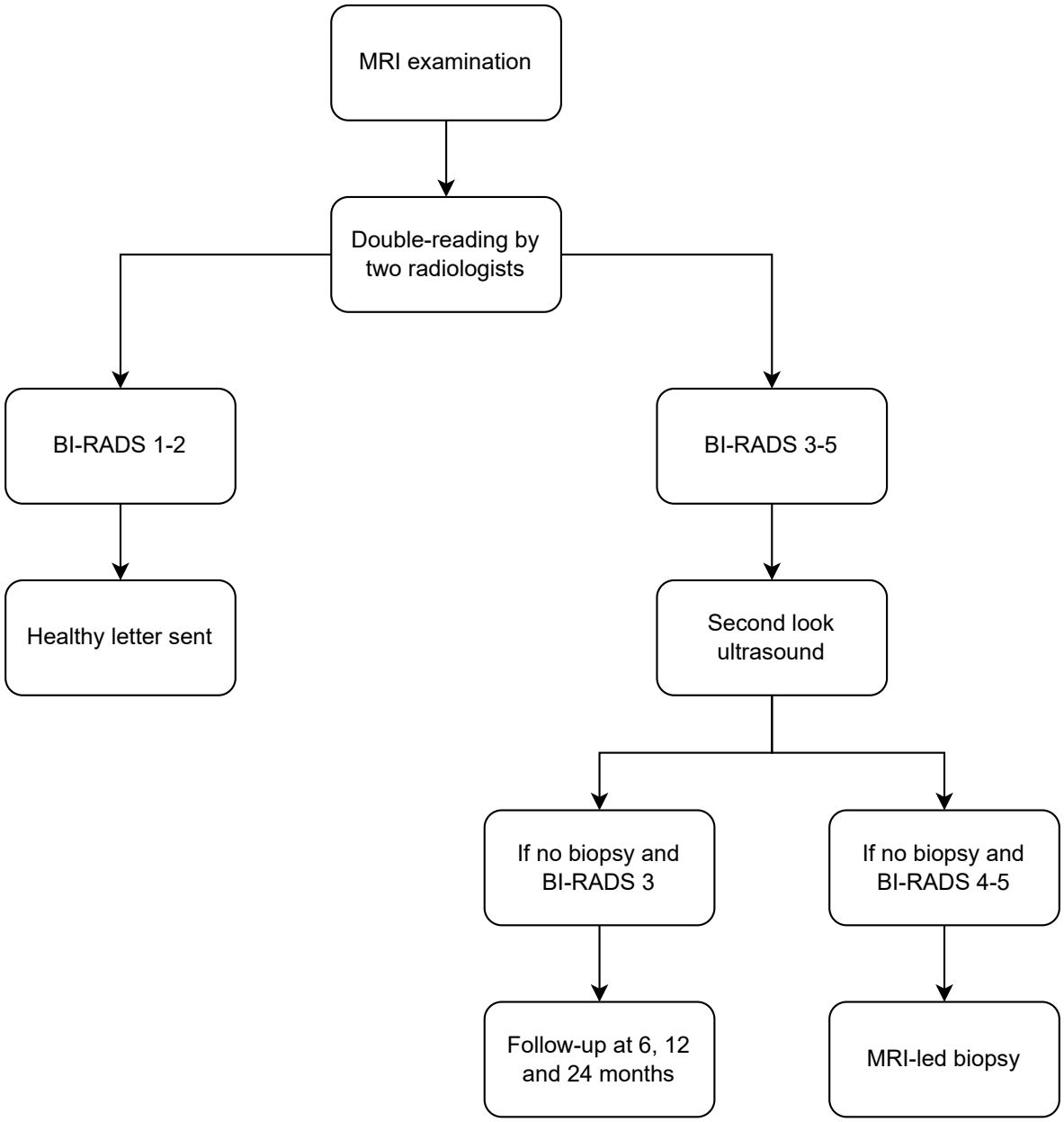
In **study I**, the data base consisted of 504,566 women invited for screening between January 1st 2008 and September 31st 2015. 83,225 women were not examined due to non-participation. 3,300 women were excluded due to unknown radiologist identity. The final study population consisted of 418,041 women and 1,186,045 screening examinations.

In **study II**, the study sample was derived from the CSAW data set used in study I. However, in this study we only included screening examinations from Karolinska University Hospital. All women between 40 to 74 years of age that were diagnosed with breast cancer between January 1st 2008 and December 31st 2015 were included in the study (n=1,253). The study excluded women that did not undergo a complete screening examination before their diagnosis, women who had a history of breast cancer and women with breast implants. We excluded all examinations with a cancer diagnosis that had a gap of more than 12 months between the examination date and diagnosis (n=419). This was because the probability of cancer being present at the time of screening was lower. We also excluded all examinations with an unknown radiologist identity (n=95). A random sample of 10,000 healthy women were included. From this sample we excluded women that had less than a 2-year cancer-free follow up (n=995), women that had examinations outside our indicated study period (n=909), women with breast implants (n=26) and also all examinations with an unknown radiologist identity (n=99). The final study population consisted of 8,066 healthy women and 739 women with breast cancer.

In **study III**, we used the same subset and inclusion criteria as in study II. Additionally, we excluded women with clinical symptoms. The final study population consisted of 8,029 healthy women and 714 women with breast cancer.

In **study IV**, we included all women undergoing mammography screening at Karolinska University Hospital from April 1, 2021 to December 31, 2022. All women with a positive screening examination were excluded from the study (n=386). We also excluded women with breast implants, breastfeeding, pregnancy and MRI contraindications such as pacemaker or other non MRI compatible implants. Women in the surveillance program for breast cancer were also excluded. All women with an AISmartDensity score above 1.97 (approximately the highest 8%) were invited to participate in the study. Of those that accepted to participate, half were randomized to supplemental MRI and the other half to be in the observational control group.

The invitation process can be seen in detail in the flowchart below (Figure 2 manuscript 4).



4.2 Register data

Sweden maintains a number of population-based registers that contain information on the populations health. In 1947 the personal identification number system was introduced which helped facilitate population-based registers. The Swedish Cancer Register was implemented in 1958. It contains data concerning type of cancer, date of diagnosis, TNM stage and histological type. It is used to monitor trends in cancer incidence, mortality and survival over time. The register is considered one of the most comprehensive and serves as a valuable resource for researchers and healthcare (174).

In studies I-III data was retrieved by linking the women using the personal identification numbers with the Screening Register at the Regional Cancer Centre Stockholm-Gotland and the breast cancer quality register which in turn receives data from the Swedish Cancer Register.

In study II, the personal numbers of all women in the subset data were linked with Karolinska University Hospital PACS (radiology image database) for the mammographic examination images, this information was later used in study III as well.

In study IV, the diagnosis of breast cancer was collected through the pathology reports stored in the medical journal system used at Karolinska University Hospital (Take care).

4.3 Readers

In study I, there were 110 interpreting radiologist. We classified radiologists into two groups: low-volume and high-volume readers. This classification was based on the number of annual screening mammograms. Radiologists who read less than 5,000 screening mammograms for at least one year during the study period were considered low-volume readers, while those who read 5,000 or more were categorized as high-volume readers. We grouped the high-volume readers into quartiles, with quartile one comprising the worst-performing readers and quartile four comprising the best-performing readers.

Studies II and III involved 25 first reader radiologists and 20 second reader radiologists. In addition, study II, involved three different commercial AI CAD algorithms for analysis of screening images.

In study IV, two radiologists evaluated all MRI examinations, while all radiologists at the Karolinska University Hospital breast radiology department conducted additional work-up for recall cases. Additionally, three different algorithms were used for analysis of screening images.

4.4 Algorithms

All deep learning algorithms included in study II were commercial. The vendors for the algorithms chose to remain anonymous, apart from algorithm 1. Details on their training can be seen in detail in table 2 below.

Table 2.

	Algorithm 1	**Algorithm 2**	**Algorithm 3**
Network architecture	ResNet-34	MobileNet	Did not disclose
Training procedure	Two-stage procedure	Two-stage procedure	Unclear
Number of training cases	72,000 cancer images, 680,000 normal images	10,000 cancer images, 229,000 normal images	6,000 cancer images, 106,000 normal images
Mammography device brand	Mostly GE	Mostly Hologic	Mostly Hologic

For study III we used the best performing algorithm (algorithm 1) from study II. For study IV, we also used algorithm 1 from study II in combination with three deep learning algorithms developed with our collaborating researchers and engineers at KTH Royal institute of technology in Stockholm. The training data was mainly derived from the CSAW dataset used in study I. The algorithms were trained to assess cancer signs, inherent risk and masking potential. Further details on the algorithms can be seen in table 3.

Table 3.

	Cancer model	**Risk model**	**Masking model**
Network architecture	Efficientnet-B3+2 residual blocks	Efficientnet-B3	ResNet-34
Training procedure	Two-stage procedure	Binary classification	Ordinal classification
Mammography device brand	Mostly Hologic	Mostly Hologic	Only Hologic

4.5 Statistical methods

All statistical analyses in this thesis were conducted using the computer software Stata, version 15.1. All statistical tests were two-sided. The level for significance was set at alpha=0.05.

Linear Regression

To model the relationship between a predictor variable and one or multiple output variables linear regression can be used. The purpose of linear regression is to find the values of the parameters that best fit the data.

In study I, using the quartile as the predictor, linear regression models were used to test for an association across quartiles of radiologists performance.

95% Confidence Interval (CI)

To estimate an unknown population parameter with a certain level of confidence 95% CI can be used. A 95% CI indicates that if the same sample were collected multiple times and a CI was calculated for each sample, we would expect the true population parameter to fall within the calculated CI 95% of the time.

The 95% CI was calculated for studies I and II.

P-value

A p-value is a probability used in statistical hypothesis testing to indicate the level of evidence against a null hypothesis. The null hypothesis is a statement of no difference between groups. A p-value is the probability of obtaining a test statistic as extreme or more extreme than the one observed, under the assumption that the null hypothesis is true. For example, if a p-value is 0.05 it means that there is a 5% chance of getting a test statistic as extreme or more extreme than the one observed, if the null hypothesis were true. If the p-value is less than the pre-specified level of significance (usually 0.05), the null hypothesis is rejected.

P-value was calculated for studies I and III.

Standard deviation

Standard deviation (SD) is a statistical measure of the dispersion of a set of data. It is a way to quantify the amount of variation or deviation from the mean of a dataset. The standard deviation is useful because it tells us how much the data deviates from the mean. A low standard deviation indicates that the data points are close to the mean, while a high standard deviation indicates that the data points are spread out over a wider range.

The standard deviation was calculated for study I.

Bootstrapping

Bootstrapping is a statistical method that involves resampling a dataset with replacement to estimate certain properties of a population, such as the mean, standard deviation or confidence intervals. In study II for example, the dataset was enriched with positive cases, so we applied stratified bootstrapping with a 14:1 ratio of healthy to diagnosed women to mimic the ratio in the source screening cohort. Stratified bootstrapping is a variation of the standard bootstrapping method, in stratified bootstrapping the data is resampled within each stratum separately, rather than sampling from the entire dataset as a whole. This ensures that the proportion of observations from each stratum in the resampled dataset is the same as in the original dataset.

Bootstrapping was used for studies I–III.

AUC

Area under the receiver operating characteristic curve (AUC), is a measure of the performance of a binary classification model. AUC ranges from 0 to 1 with a value of 1 indicating a perfect model and a value of 0.5 indicating a model that performs no better than random. AUC can be useful in assessing the performance of a model because it is independent of the classification threshold and can provide an overall measure of the model´s accuracy. It also allows for the comparison of different models and can be used to select the best model for a given dataset.

AUC was calculated for study II using the DeLong method and the AUC CIs were estimated by the sandwich variance estimator.

DeLong method

The DeLong method is a statistical method for comparing the AUC of two different diagnostic tests.

The DeLong method was used for study II.

IQR

Interquartile range (IQR) is a statistical measure used to describe the spread of a dataset. The IQR is defined as the difference between the 75^{th} and 25^{th} percentiles of the dataset.

IQR was used for studies II, III and IV.

5 Results

5.1 Study I

In this study we analyzed 1,186,045 screening mammograms for 418,041 women. Among these mammograms, 972,899 were assessed by high volume readers, while the remaining 213,146 were assessed by low-volume readers. A total of 4,723 women were diagnosed with breast cancer either at screening or within a period of 12 months. The mean age at screening was 54 years (SD 9.5), and the mean age at diagnosis was 59 years (SD 10.1). The study included 24 high-volume readers and 86 low-volume readers. While the sensitivity and specificity measures were similar for both groups, there was a notable discrepancy between Q1 (quartile 1) and Q4 (quartile 4), especially regarding sensitivity.

The below table reports the first-reader performance measures (Table 2 article 1).

Table 2: Screening Performance Benchmarks Overall and for Each Quartile of High-Volume Readers

Parameter	All Interpreting Radiologists (n = 110)	Low-Volume Readers* (n = 86)	High-Volume Readers* (n = 24)	High-Volume Reader Performance			
				Q1 (n = 6)	Q2 (n = 6)	Q3 (n = 6)	Q4 (n = 6)
No. of examinations	1 186 045	213 146	972 899	…	…	…	…
Sensitivity (%)	73	75	73	63 (53–67)	68 (66–75)	78 (75–79)	84 (79–86)
Specificity (%)	96	96	96	95 (93–96)	96 (96–96)	97 (96–97)	98 (98–99)
AIR	39	35	40	56 (50–72)	46 (40–49)	39 (34–40)	24 (15–31)
CDR	3.0	2.4	3.1	2.2 (1.7–2.5)	2.9 (2.6–3.3)	3.6 (3.7–3.6)	3.9 (3.7–4.7)
FNR (%)	29	42	26	38 (34–47)	31 (25–33)	22 (21–23)	16 (14–21)
Accuracy (%)	96	96	96	95 (93–95)	96 (95–96)	97 (96–98)	98 (98–99)
Positive predictive value (%)	8	7	8	6 (4–7)	7 (7–8)	10 (8–12)	12 (12–14)

Note.—Unless otherwise specified, data are means, with the range in parentheses. Mean values represent averages of the radiologist-level metrics. Q1 is the lowest quartile, and Q4 is the highest. AIR = abnormal interpretation rate (per 1000 examinations), CDR = cancer detection rate (per 1000 examinations), FNR = false-negative rate (per cancer diagnosed within 12 months).

* The distinction was based on the maximum number of annual examinations, with high-volume readers performing at least 5000 in any year.

We also analyzed the screening outcomes according to sensitivity quartile of high-volume readers based on 10,000 screening mammograms per quartile. The most sensitive radiologists (Q4) diagnosed more cancers than the least sensitive radiologists (Q1) however, the abnormal interpretation rate was substantially higher for Q4 compared to Q1. For Q1, 14 cancers were missed and for Q4, 7 cancers were missed.

The screening outcomes per 10,000 women for each quartile of high-volume reader according to sensitivity can be seen in the below table (Table 3 article 1).

Table 3: Screening Outcomes per 10 000 Women Screened for Each Quartile of High-Volume Reader according to Sensitivity

	High-Volume Readers Divided into Quartiles according to Sensitivity*			
Outcome	Q1 (n = 6)	Q2 (n = 6)	Q3 (n = 6)	Q4 (n = 6)
All screened	10,000	10,000	10,000	10,000
Cancer	38	41	45	45
Healthy	9962	9959	9955	9955
Abnormal interpretation	281	404	478	497
Cancer	24	31	35	37
Healthy	257	373	443	460
PPV (%)	8.6	7.6	7.3	7.5
Normal interpretation	9719	9597	9523	9502
Cancer	14	11	10	7
Healthy	9705	9586	9513	9495

Note.—Unless otherwise specified, data are numbers of examinations. Q1 is the lowest quartile, and Q4 is the highest. PPV = positive predictive value.

* The distinction was based on the maximum number of annual examinations, with high-volume readers performing at least 5000 in any year.

We also examined first reader sensitivity for each tumor subgroup. The molecular subtypes were defined according to the St. Gallen consensus (64). Analysis by quartile of high-volume readers revealed that the sensitivities for the most sensitive high-volume readers (Q4) were 85% for invasive cancers and 99% for in situ cancers only (high grade). The biggest variation in performance was observed for the basal molecular subtype, where the sensitivity for the least sensitive quartile (Q1) was 53%, while the sensitivity for the most sensitive quartile (Q4) was 89%.

Sensitivity for all tumor subgroups can be seen in the table below (Table 4 article 1).

Table 4: First-Reader Sensitivity according to Tumor Characteristics

		Sensitivity						
				High-Volume Readers Divided into Quartiles according to Overall Sensitivity				
Characteristic	No. of Tumors	Low-Volume Readers*	High-Volume Readers*	Q1	Q2	Q3	Q4	P Value†
Invasive tumor size‡								
Minimal (≤10 mm)	888	81	80	64	82	83	88	<.001
Small (11–19 mm)	1807	74	78	67	77	82	85	<.001
Large (≥20 mm)	1014	74	72	64	61	77	84	<.001
Invasive tumor histologic type								
Ductal§	3189	76	77	67	78	82	85	<.001
Lobular	507	75	73	63	76	76	84	<.001
Other	254	77	76	62	80	82	86	<.001
Invasiveness								
Invasive	3965	76	76	67	78	81	85	<.001
In situ only, any grade	698	84	83	75	85	84	93	<.001
In situ only, high grade	331	90	87	75	93	90	99	<.001
Molecular subtype								
Luminal A	2916	76	77	67	79	82	85	<.001
Luminal B	296	72	73	62	73	75	88	<.001
Her2 overexpressing	119	68	76	67	78	82	82	<.001
Basal	212	66	69	53	66	74	89	<.001

Note.—Unless otherwise specified, data are percentages. Q1 is the lowest quartile, and Q4 is the highest.

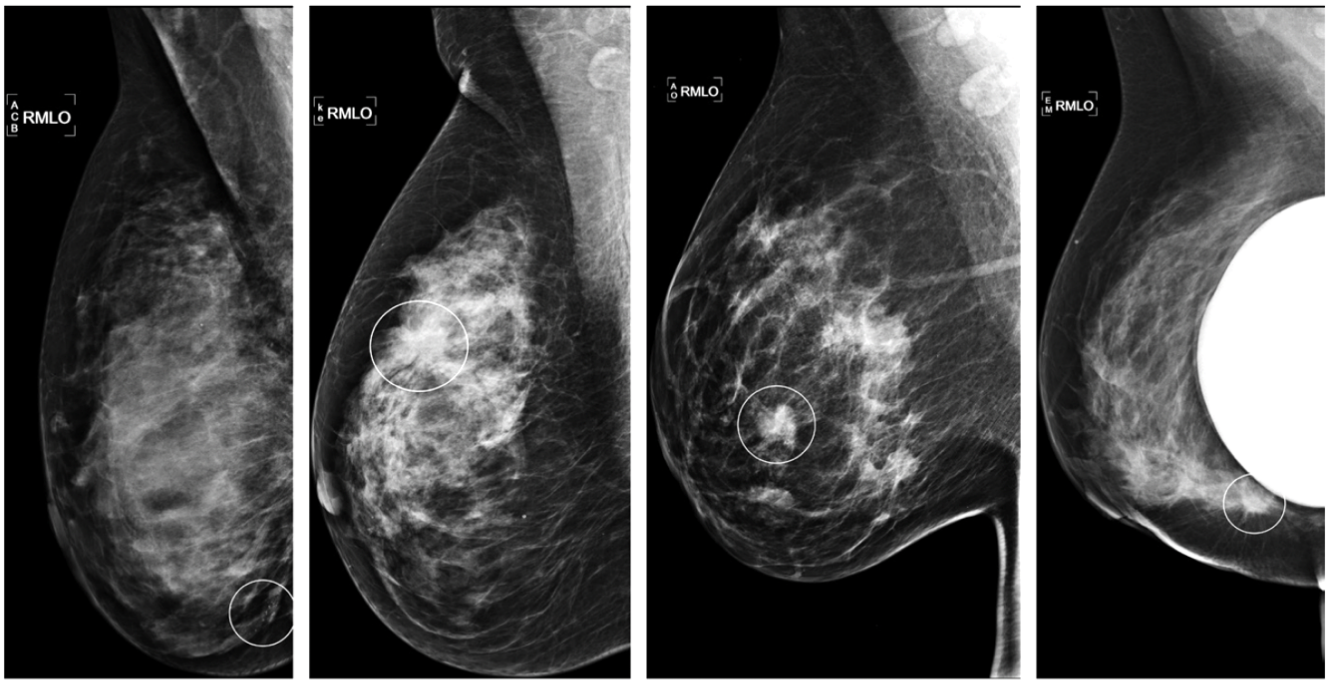
* The distinction was based on the maximum number of annual examinations, with high-volume readers performing at least 5000 in any year.

† P value for the association across quartiles.

‡ Tumors subject to neoadjuvant therapy were not included.

§ Ductal includes tumors that are only ductal, as well as mixed tumors with a ductal component.

Mammograms that were assessed as true-positive or false-negative by the first reader can be seen below (Figure 2 article 1).



a. b. c. d.

Figure 2: Mammograms in which the first reader made a true-positive or false-negative assessment. **(a)** True-positive right mediolateral oblique (RMLO) mammogram in a 49-year-old woman with diagnosis of invasive 4-mm ductal breast cancer within 13 days of screening shows a cluster of calcifications. **(b)** True-positive RMLO mammogram in a 66-year-old woman with diagnosis of invasive 21-mm ductal breast cancer within 12 days of screening shows a spiculated mass. **(c)** False-negative RMLO mammogram in a 52-year-old woman with diagnosis of invasive 23-mm ductal breast cancer within 9 days of screening shows an indistinct mass. **(d)** False-negative RMLO mammogram in a 44-year-old woman diagnosis of invasive 36-mm ductal breast cancer within 26 days of screening shows a spiculated mass.

5.2 Study II

In this study we incorporated 8,805 women and screening examinations, of which 739 women had been diagnosed with breast cancer and 8,066 women were included as healthy controls. The median age at screening was 54.5 years (IQR, 47.4-63.5 years), and the median age at diagnosis was 59.8 years (IQR, 49.8-5.8).

The study examined the AUC of each algorithm for detecting cancer, both overall and within subgroups. Overall, the AUC was 0.956 (95% CI, 0.948-0.965) for AI algorithm 1 (AI-1), 0.922 (95% CI, 0.910-0.934) for AI-2, and 0.920 (95% CI, 0.909-0.931) for AI-3. There was a statistically significant difference (P<.001) between AI-1 and the other two AI algorithms (AI-2 and AI-3). However, there was no statistical significance for the difference between AI-2 and AI-3 (P=.68). Additionally, for all the analyzed subgroups, AI-1 displayed a significantly higher AUC compared to AI-2 and AI-3.

We also calculated the AUC for younger and older women in addition to lower and higher breast density. For the best performing algorithm we observed that the AUC was lower for younger women in comparison to older; 0.974 for women 55 years or older and 0.925 for women younger than 55 years. For density the corresponding numbers were 0.933 for mammograms with high percent density and 0.976 for mammograms with low percent density.

The AUC's for all algorithms overall and by subgroups can be seen in the table below (Table 1 article 2).

Table 1. Area Under the Receiver Operating Characteristic Curve for the 3 Artificial Intelligence Algorithms

	AUC (95% CI)[a]		
Group (n = 8805)	Algorithm 1	Algorithm 2	Algorithm 3
Overall	0.956 (0.948-0.965)	0.92 (0.910-0.934)	0.92 (0.909-0.931)
By age, women, y			
Younger (<55)	0.925 (0.906-0.944)	0.882 (0.856-0.907)	0.889 (0.867-0.912)
Older (≥ 55)	0.974 (0.966-0.982)	0.943 (0.932-0.954)	0.938 (0.927-0.949)
By mammographic density[b]			
Dense area[c]			
Low	0.973 (0.964-0.981)	0.945 (0.932-0.959)	0.940 (0.926-0.954)
High	0.938 (0.923-0.954)	0.899 (0.879-0.918)	0.900 (0.882-0.917)
% Density[c]			
Low	0.976 (0.968-0.983)	0.954 (0.943-0.966)	0.950 (0.939-0.961)
High	0.933 (0.917-0.950)	0.886 (0.865-0.908)	0.886 (0.867-0.906)
By cancer detection mode			
Screen	0.984 (0.979-0.989)	0.959 (0.951-0.967)	0.952 (0.944-0.960)
Clinical	0.810 (0.767-0.852)	0.728 (0.677-0.779)	0.744 (0.696-0.792)

Abbreviation: AUC, area under the receiver operating characteristic curve.

[a] Test: Algorithm 1 has a higher AUC than the other 2 algorithms overall and for all subgroups ($P < .001$).

[b] Examination mean of all 4 views.

[c] Low represents below median; high, above median.

Additionally, we compared the performance benchmarks of all algorithms with that of the radiologists. After bootstrapping the total simulated screening population consisted of 112,924 examinations for healthy women and 739 examinations for women with a breast cancer diagnosis. The sensitivities were 81.9 % for AI-1, 77.4% for the first reader, 80.1% for the second reader and 98.5% for the consensus discussion.

There was a statistically significant difference (P=0.03%) in sensitivity between AI-1 and the first reader (P=0.03%), while no statistical significance was observed in the comparison between AI-1 and the second reader (P=.40) or the consensus discussion (P=.11). The specificity of the AI algorithms was pre-defined to match that of the first reader, hence it could not be compared.

In the below table (Table 2 article 2) the screening performance benchmarks for the comparison between the AI algorithms and the radiologists can be seen in more detail.

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

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

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

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

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

Additionally, we simulated three different scenarios by combining the binary decisions of the three AI algorithms and the readers. For the first reader, the addition of AI-1 resulted in a 15% relative increase in cancer detection, whereas the addition of the second reader lead to a 12% relative increase. Furthermore, the incorporation of AI-1 resulted in a 78% relative increase in abnormal interpretations, while the addition of the second reader led to a 24% relative increase.

The results of the simulated scenarios can be seen in more detail in the table below (Table 3 article 2).

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

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

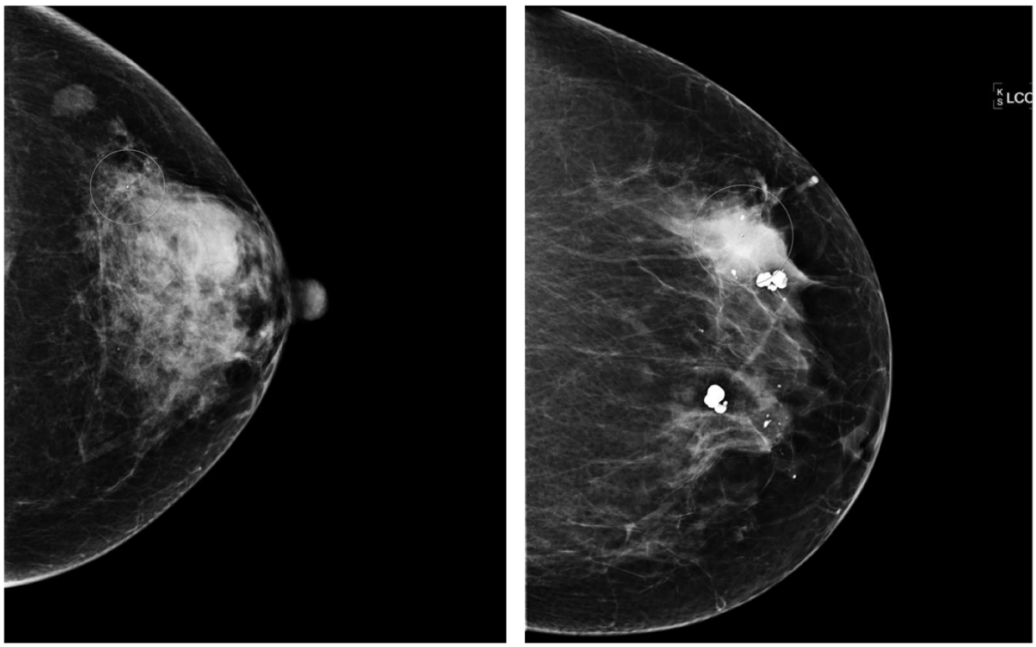
Abbreviation: NA, not applicable.

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

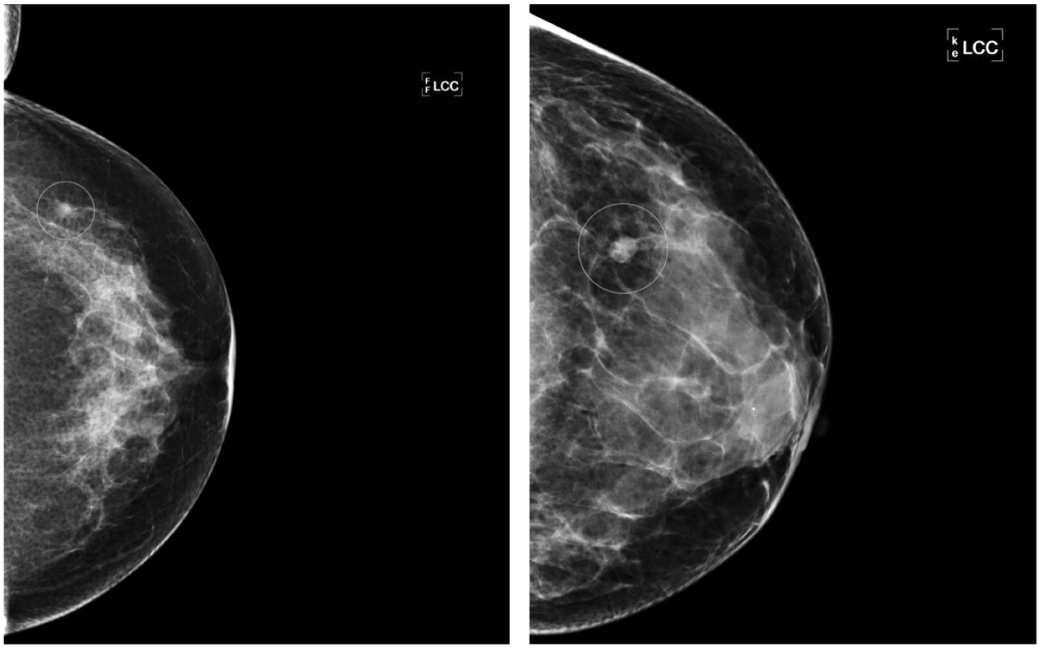
[b] Actual screen-detected cancer (n = 618); actual clinically detected cancer (n = 121).

The images below are examples of mammograms with cancers either missed by both radiologists but identified by AI or missed by all three AI but identified by radiologists (eFigure 2 & eFigure 3 article 2).

eFigure 2. Examples of Mammograms of Cancer Identified by AI CAD but Missed by Both Radiologists



eFigure 3. Examples of Mammograms of Cancer Identified by Radiologists but Missed by All Three AI CAD



5.3 Study III

In this study we included 714 women who were diagnosed with breast cancer and a random sample of 8,029 healthy controls. After bootstrapping, the total simulated screening population comprised 113,120 screening examinations from 8,743 women. The median age at screening was 54.1 (IQR=47-63), and the median age at diagnosis was 60.2 (IQR=50-66).

We analyzed the distribution of false assessments between AI and the radiologists and observed that for AI, RAD 1 (first reader) and RAD 2 (second reader), the FN (false negative) assessments were more frequent in high-density than low-density women. The FN assessments were distributed between low- and high-density women, with 53 (42%) and 72 (58%) for AI CAD, 59 (36%) and 104 (64%) for RAD 1, and 47 (36%) and 84 (64%) for RAD 2. Regarding age, the distribution of FN assessments between younger and older women were 69 (55.2%) and 56 (44.8%) for AI CAD, 88 (54%) and 75 (46%) for RAD 1 and 76 (58%) and 55 (42%) for RAD 2. In contrast to AI, the FP assessments for RAD 1 were more common in younger women (p=<0.001). However, AI had more FP assessments for older women compared to both radiologists (p=<0.001).

Additionally, we conducted a simulation of a double-reading scenario by combining AI with RAD 1 and AI with RAD 2, and compared the results with the actual double-reading combination of RAD 1 and RAD 2. The addition of AI to any radiologists resulted in a significantly higher number of false positive assessments (p=<0.001); 6790 (6%) for AI + RAD 1; 6,202 (5.5%) for AI + RAD 2 and 4,382 (3.9%) for RAD 1 + RAD 2. The corresponding numbers for false negative assessments were 83 (11.6%); 75 (10.5%) and 98 (13.7%), respectively. The largest increase in false positive assessments was noted for older (>=55) women when combining AI with radiologists compared to radiologists only, for AI + RAD 1 the increase was 89% and for AI + RAD 2 the increase was 78% (p=<0.001).

The results from the simulated double-reading can be seen in more detail in the table below (Table 3 manuscript 3).

Table 3. Double-reading. Errors in assessments across age, density and tumor characteristics

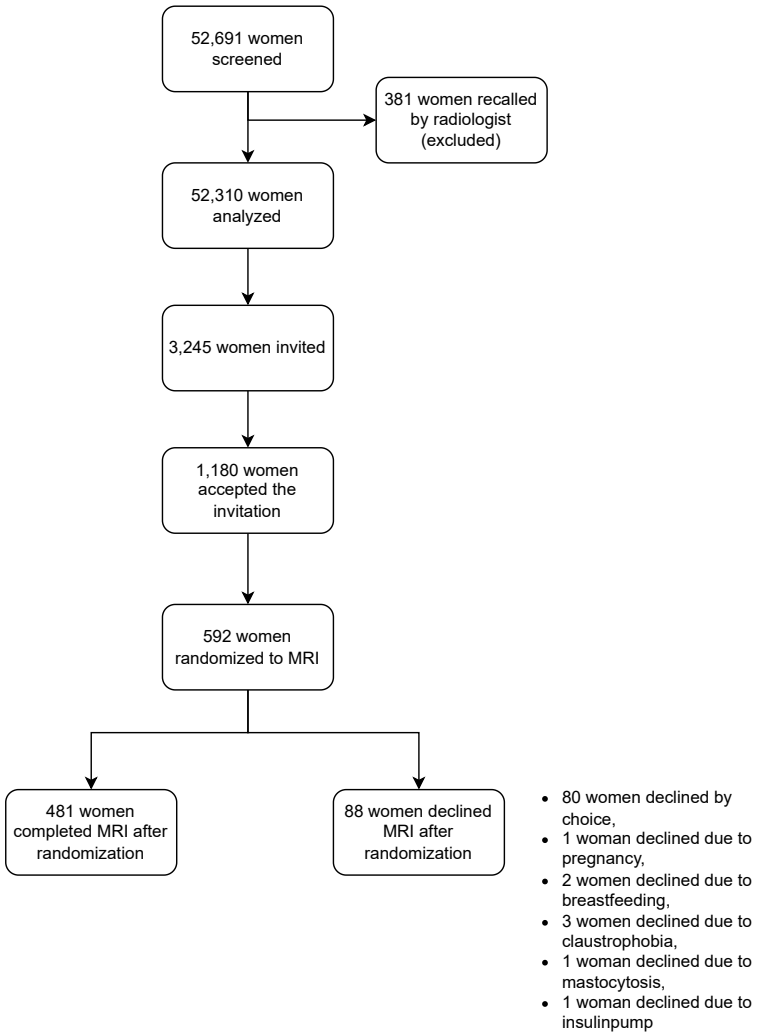
	RAD 1 + RAD 2 FALSE POSITIVE		AI+RAD 1 FALSE POSITIVE					AI+RAD 2 FALSE POSITIVE				
n and %												
Total n cases (n=113,120)	**Total n**	**Absolute % *****	**Total n**	**Absolute %**	**Change n****	**Change %****	***P*-value******	**Total n**	**Absolute %**	**Change n****	**Change %****	***P*-value******
All false positive	**4,382**	**3.9%**	**6,790**	**6.0%**	**+2,408**	**+55%**	**<0.001**	**6,202**	**5.5%**	**+1,820**	+42%	<0.001
Age<55	2,828	2.5%	3,850	3.4%	+1,022	+36%	<0.001	3,430	3.0%	+602	+21%	<0.001
Age>=55	1,554	1.4%	2,940	2.6%	+1,386	+89%	<0.001	2,772	2.5%	+1,218	+78%	<0.001
Low-density	1,988	1.8%	3,136	2.8%	+1,148	+58%	<0.001	2,814	2.5%	+826	+42%	<0.001
High-density	2,394	2.1%	3,654	3.2%	+1,260	+53%	<0.001	3,388	3.0%	+994	+42%	<0.001
	RAD 1 +RAD 2 FALSE NEGATIVE		**AI+RAD 1 FALSE NEGATIVE**					**AI+RAD 2 FALSE NEGATIVE**				
All false negative Cancer(n=714)	**98**	**13.7%**	**83**	**11.6%**	**-15**	**-15%**	**0.233**	**75**	**10.5%**	**-23**	-24%	0.062
Age <55	63	8.8%	56	7.8%	-7	-11%	0.464	51	7.1%	-12	-19%	0.202
Age >=55	35	4.9%	27	3.8%	-8	-23%	0.293	24	3.4%	-11	-31%	0.139
Low-density	29	4.1%	24	3.4%	-5	-17%	0.475	22	3.1%	-7	-24%	0.308
High-density	69	9.7%	59	8.3%	-10	-15%	0.331	53	7.4%	-16	-23%	0.113
Histology* (n=616)	**86**	(12 missing)										
Ductal	74	12.0%	61	9.9%	-13	-18%	0.229	56	9.1%	-18	-24%	0.090
Lobular	10	1.6%	8	1.3%	-2	-20%	0.616	6	1.0%	-4	-40%	0.291
Other	2	0.3%	1	0.2%	-1	-50%	0.556	2	0.3%	0	0%	1.000
Invasiveness (n=700)	**91**	(7 missing)										
Invasive component	86	12.3%	70	10.0%	-16	-19%	0.171	64	9.1%	-22	-26%	0.055
In situ only	5	0.7%	5	0.7%	0	0%	1.000	5	2.3%	0	0%	1.000
Molecular subtype (n=573)	**75**	(23 missing)										
Luminal A	53	9.2%	45	7.9%	-8	-15%	0.393	37	6.5%	-16	-30%	0.076
Luminal B	9	1.6%	6	1.0%	-3	-33%	0.401	6	1.0%	-3	-33%	0.401
Her2-overexpressing	4	0.7%	4	0.7%	0	0%	1.000	4	0.7%	0	0%	1.000
Basal	9	1.6%	7	1.2%	-2	-22%	0.578	8	1.4%	-1	-11%	0.786

*) Ductal includes tumors that are only ductal as well as mixed tumors with a ductal component
**) Change is the total number of false assessments for AI + RAD divided by the total number of false assessments for RAD 1 + RAD 2
***) Absolute % is the number of false assessments divided by all examinations
****) P-value for the absolute difference between AI+RAD and RAD 1+RAD 2

5.4 Study IV

During the indicated study period, we analyzed 52,310 examinations of which 3,668 (7%) women had above AISmartDensity threshold and 3,245 (6.2%) were invited to participate. Of those invited, 1,180 consented to participate, and 592 women were randomized to complementary MRI screening. The median age of women invited to the study was 58 (IQR;49-66).

The study population can be seen in detail in the flowchart below (Figure 1 manuscript 4).

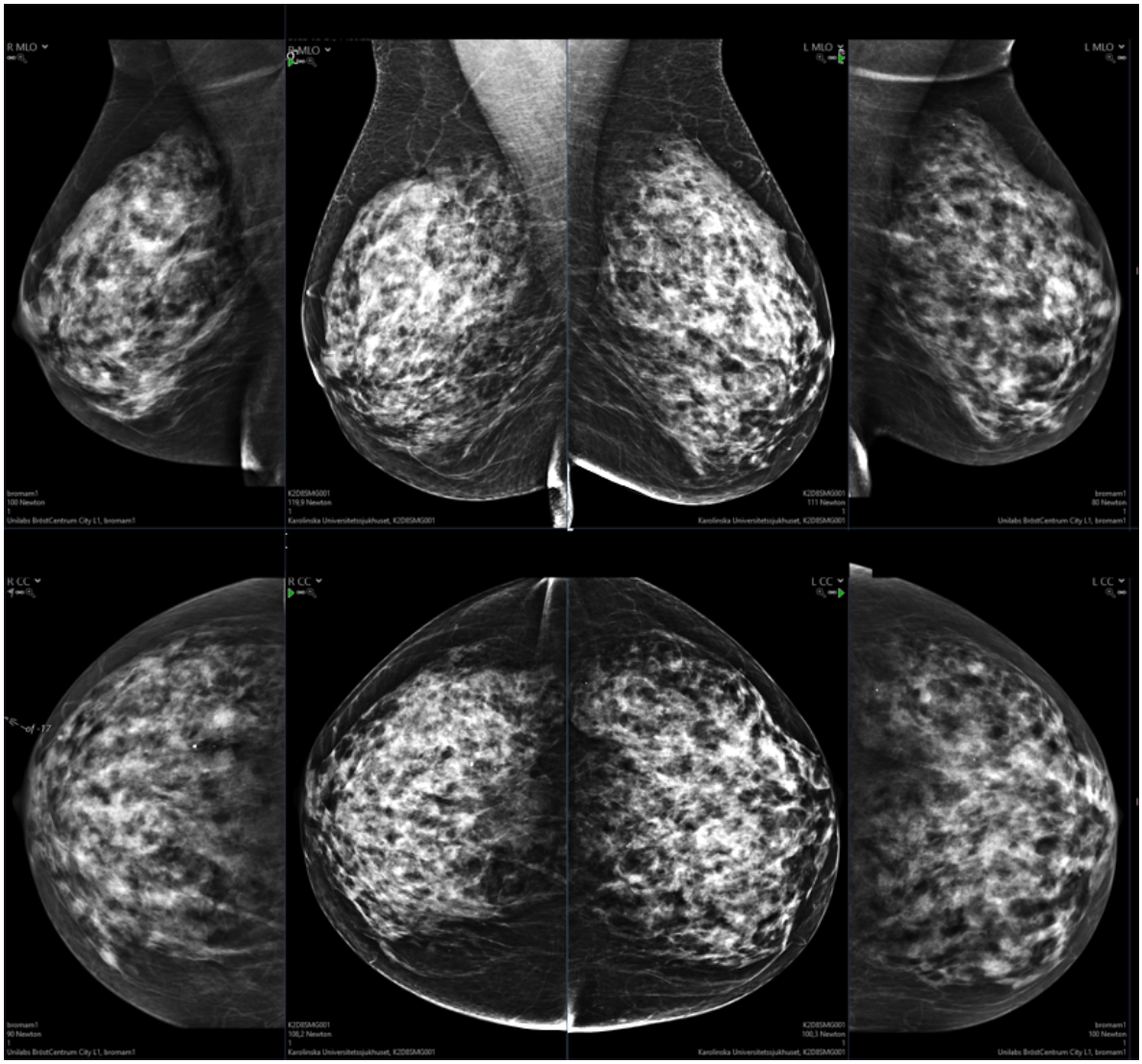


In this interim report we evaluated the outcomes of MRI and biopsy for the 481 (81%) women who underwent MRI. The median AISmartDensity score was 2.39 (IQR;2.15-2.87). Out of all completed MRI examinations, 399 (83%) were categorized as BI-RADS 1-2, 48 (10%) had a BI-RADS 3 finding, 20 (4%) had a BI-RADS 4 finding and 14 (3%) had a BI-RADS 5 finding. Biopsies were guided by ultrasound for 54 women, by MRI for 4 women and by stereotactic biopsy for 1 woman. Among women who underwent biopsy, 48% had breast cancer. CDR was 58.2 per 1,000 MRI examinations. For all BI-RADS 3-5 examinations, PPV and FPR were 34% and 13%, respectively.

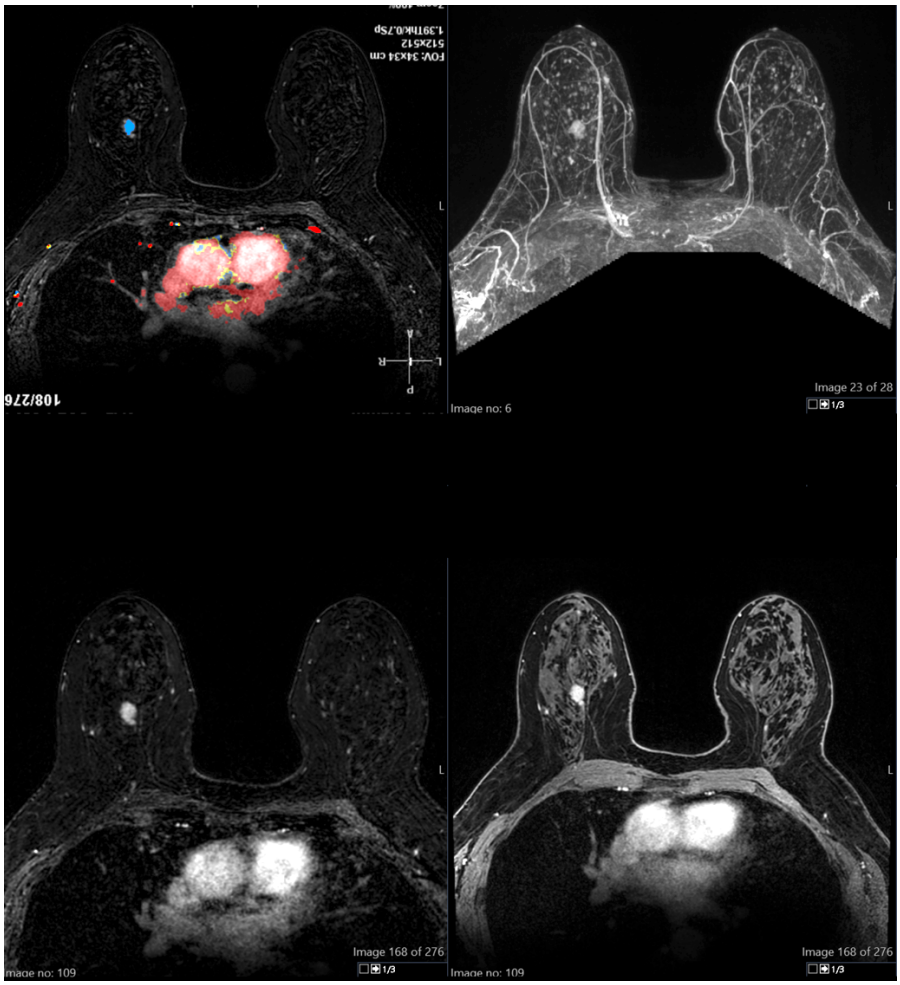
The MRI and biopsy outcome can be seen in more detail in the table below (Table 2 manuscript 4).

Table 2. MRI and Biopsy outcome							
	Total, n	**Not Biopsied**	**Biopsied**	**Benign* (n)**	**Cancer * (n)**	**CDR****	**PPV of BIRADS 3-5**
BI-RADS 1	304	N/A	N/A	N/A	N/A	N/A	N/A
BI-RADS 2	95	N/A	N/A	N/A	N/A	N/A	N/A
BI-RADS 3	48	24	24	19	5	N/A	10%
BI-RADS 4	20	1	19	8	11	N/A	55%
BI-RADS 5	14	0	14	2	12	N/A	86%
All, n	481	25	59	31	28	58.2	34%
*) Biopsy verified							
**) CDR = Cancer Detection Rate (per 1000 examinations).							

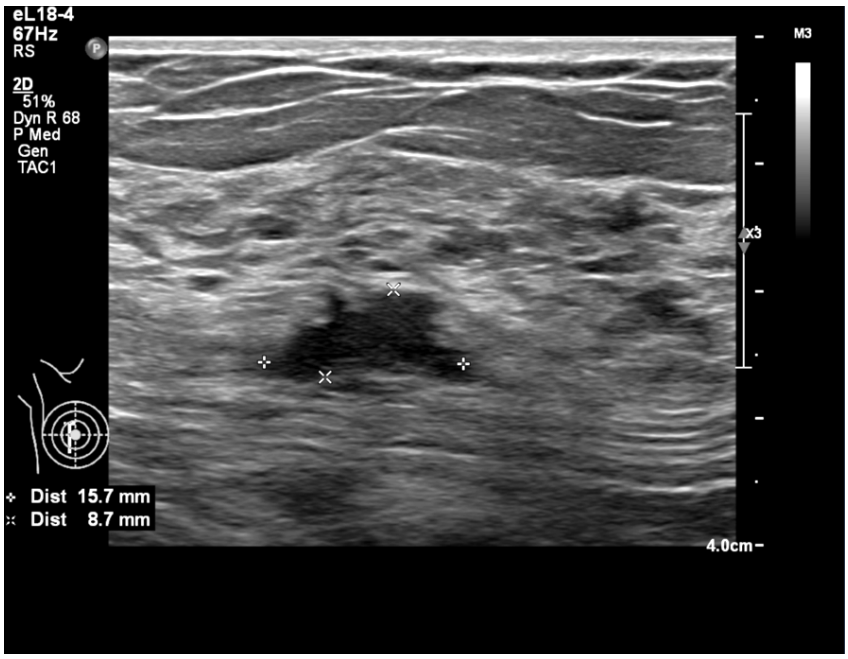
Below you can see a case example from the study. A 72-year old woman that underwent screening mammography. The images in the middle are from the current mammographic examination and the images outwards are from the previous examination two years ago. There were no signs of malignancy on her examination and she was declared healthy. She scored above threshold AISmartDensity and was invited to the study and randomized to MRI.



On the MRI examination below you can see a BI-RADS 3 finding in the right breast.



The second look ultrasound revealed a suspicious lesion that was biopsied.



She underwent surgical excision and the pathology report revealed an invasive cancer with mucinous differentiation and DCIS.

6 Discussion

6.1 Study I

In this study we explore the performance benchmarks for first readers of screening mammograms based on a dataset of more than 1 million screening examinations.

In our study the sensitivity and specificity were similar between low- and high-volume readers. These results suggest that there is no significant difference in sensitivity and specificity between low- and high-volume readers. This means that readers who interpret fewer mammograms perform just as well as those who interpret a larger volume of mammograms in terms of detecting breast cancers and avoiding false positives. Additionally, we observed significant variations in sensitivity among individual high-volume readers. Interestingly, this indicates that the experience level of radiologists may not be the only factor that affects their accuracy in interpreting mammograms. Previous studies have reported contradictory data regarding the relationship between experience level and accuracy in interpreting mammograms (175, 176, 177), this underscores the need for further research into this topic. It is possible that other factors, such as the complexity of the cases presented or the training of the radiologists, may influence the relationship between experience level and accuracy in mammogram interpretation.

Another key finding was that radiologists who operated at higher sensitivity levels had a lower rate of false-negative screenings by the first reader, which could indicate a reduced incidence of interval cancers. However, these same radiologists also had a higher incidence of abnormal interpretations, which could cause additional recalls. These findings are consistent with a prior study by Burnside et al. (178) which establishes a correlation between higher recall rates and lower interval cancer rates. This highlights the importance of a balance between sensitivity and specificity when interpreting mammograms to optimize the detection of breast cancer while minimizing the number of unnecessary recalls.

Another important finding was the difference in sensitivity observed among radiologist for different tumor subgroups. We found that the most sensitive radiologists had a 99% sensitivity level for high-grade in situ cancers, in contrast to 85% for invasive cancers. The greater sensitivity for in situ cancers could potentially be explained by their mammographic appearance of calcifications. Calcifications are generally visible on a mammogram and often lack accompanying clinical symptoms, making them less likely to be detected by the woman (89, 179, 180).

Finally, our study found that the largest difference in performance among radiologists was observed for the basal molecular subtype, with the most sensitive quartile

detecting 89% and the least sensitive quartile detecting only 53%. This difference in sensitivity could be partly explained by the fact that basal cancers often have benign mammographic findings that make them easier to miss (181).

6.2 Study II

In this study we evaluated the performance of three commercially available AI systems for screening mammography and compared them to the performance of the radiologists.

We observed a difference in the AUC among the three AI systems ranging from 0.920 to 0.956. The difference in AUC among the three AI systems suggests that some AI algorithms may perform better than others. We also found that the best AI algorithm reached, and in some comparisons surpassed, the performance level of radiologists in assessing screening mammograms with a sensitivity level of 81.9%. Additionally, we found that combining the first reader with the best algorithm identified more cancer cases than combing the first and second readers.

A subgroup analysis revealed a decreased performance for younger women and women with higher mammographic breast density. This is consistent with previous studies that have concluded that there is a decreased mammographic sensitivity for younger women and women with higher breast density (26, 182, 183, 184). These results highlight the need for improvement in the performance of the AI systems in these subgroups.

The superior performance of AI-1 compared to the other two algorithms may be due to factors such as its larger training dataset, pixel-level annotations, higher capacity backbone and data augmentation techniques. Additionally, AI-1 was very robust as it performed well on images acquired from different equipment (it was trained on images from GE equipment and the images in the study was acquired on Hologic equipment) and on a different population (South Korean women vs Scandinavian women in the study). The superior performance of AI-1 indicate that for developing algorithms, the volume of cases may hold greater significance than the diversity of vendors or patient populations used in the training of the algorithm.

Overall, the results indicate that AI systems have the potential to improve the accuracy and efficiency of screening mammography, but further research is needed to optimize the performance and address potential limitations.

Freeman et al. (160) conducted a systematic review of 12 studies that focused on the use of AI for image analysis in breast cancer. One of the studies included in the review was our study, study II. The authors noted that there was a lot of bias present in the reviewed studies, which could affect the generalizability of the results. Their conclusion was that the current evidence is not yet sufficient to determine the accuracy of AI in screening. They suggested that further prospective studies are needed to evaluate the effectiveness of AI in breast cancer screening programs. For our study, the authors noted bias in the selection of randomly chosen controls. However, random selection processes are designed to avoid bias and was partly why we chose randomization. Furthermore, the authors of the review indicated that the

applicability of the studies to European or UK breast cancer programs was low. However, our study was based on a true population-based screening cohort and also compared the performance of the AI systems with the decision from the consensus discussion making it fully applicable to European breast cancer programs. Nonetheless, I fully agree with their statement regarding the need for further prospective studies to evaluate the potential benefits and limitations of AI in clinical practice.

6.3 Study III

In this study we evaluated the differences and similarities in false assessments by an AI algorithm and radiologists in mammography screening, with a focus on the impact of age and breast density.

We found that a younger patient age group caused more FP assessments for radiologists than for AI, but there was no associated decrease in FN assessments. This is consistent with prior studies that have reported higher FP rates among younger women, which decreased with increasing age (185, 186).

On the contrary, breast density did not cause a difference in FP assessments between radiologists and AI, but radiologists made more FN mistakes for high-density women compared to AI. This is in line with previous studies that demonstrated the masking effect of dense breast tissue, making it harder for radiologists to detect cancers in these women (26, 40, 41, 42, 43, 44, 45).

We also found that adding AI as a second reader in a double-reading simulation decreased the FN assessments for high-density women, leading to an improvement in diagnostic performance of assessments in dense breasts. This was most notable for high-density older women. These results anticipate that radiologists might be cautious when assessing images for younger patients, leading to both a higher FP rate for younger high-density women and a higher FN rate for older high-density women.

Overall, the study highlights the potential benefits of incorporating AI as a second reader in mammography screening, particularly for high-density women.

The role of breast density in optimizing breast cancer screening is increasingly becoming an important topic of discussion and the results of this study emphasize the importance of elevating that discussion. In addition to making it more difficult for radiologists to identify tumors on mammography, dense breast tissue serves as an independent risk factor for breast cancer (10, 26, 40, 187). In the United States, some states have been reporting breast density in mammography reports over the past years, while others have not. However, the FDA recently (2023) updated its regulations, mandating breast density reporting in all US states. Since 2022, EUSOBI (European Society Of Breast Imaging), also recommend informing women about their breast density and its diagnostic and prognostic implications. Additionally, they recommend supplemental screening for women with extremely dense breasts, preferably with MRI. By increasing awareness of breast density and developing individualized screening plans, early detection of breast cancer and more effective treatment can be provided.

6.4 Study IV

In this study we present interim results for ScreenTrust MRI, a prospective clinical trial using AI to select women for supplemental MRI examinations in breast cancer screening.

We found that supplemental MRI screening after a negative screening mammogram in women with above threshold AISmartDensity resulted in a CDR of 58.2 cancers per 1,000 examinations. Many studies have explored supplemental screening methods to improve the cancer detection rate in women with dense breasts.

Recently, a meta-analysis conducted by Hussein et al. (133) aimed to evaluate different supplemental screening methods for women with dense breasts and negative mammogram results. Out of 22 studies, only three studies covered breast MRI. The CDR for MRI was 25.7 per 1,000 examinations. The three MRI studies evaluated had different inclusion criteria, age range and study designs than in our study, this could account for the differences in CDR. The CDR for the study by Bakker et al. was 16.5 per 1,000 examinations, whereas the studies by Kuhl et al. and Chen et al. reported rates of 28.9 per 1,000 examinations and 33.5 per 1,000 examinations, respectively.

In our study the CDR was higher than that reported in all the MRI studies included in the meta-analysis. However, among the studies analyzed Chen et al. (188) had the highest CDR, although the indication for MRI in their study was unclear. Additionally, the ethnic composition of the study differed from our study. Also, it is important to note that China has a lower screening rate for women compared to Western countries (189), and Asian women typically have more dense breast tissue than European women and an earlier age of breast cancer onset (190, 191). These factors are all associated with advanced stage breast cancer and could explain the higher CDR in the study by Chen et al. compared to the other two studies. When comparing the results of our study to the results of the DENSE trial by Bakker et al. we note some similarities but also certain differences. The Dutch screening program is similar to the Swedish program with a two view mammography acquired biannually and assessed by two independent radiologists. However, the difference is the age of which women are invited, the Dutch program invites women 50 to 75 years of age unlike the Swedish program that invites women 40 to 74 years of age.

While mammography remains the standard screening tool for breast cancer, it has its limitations, particularly in women with dense breast tissue (10, 40). MRI has demonstrated a greater sensitivity for detecting breast cancer in this patient group (132, 133). However, replacing mammography with MRI for all women in breast cancer screening is not feasible due to MRI being more expensive and time consuming compared to mammography. Therefore, it is essential to identify the women who would benefit the most from a supplemental MRI examination. While the recommendation by EUSOBI to offer MRI to all women with extremely dense breasts seems promising, our results suggest that using artificial intelligence to select women for supplemental MRI screening after negative mammography yields a higher cancer detection rate than using breast density as a selection criterion.

7 Conclusions

In this thesis we have shown that the incorporation of AI in mammography screening can improve the accuracy and efficiency of breast cancer detection. So far, the breast cancer screening process has remained uniform for all women in screening despite the growing scientific evidence regarding the promising use of AI and the importance of additional imaging for women with dense breasts. The incorporation of AI into breast cancer screening has great possibilities and can facilitate tailored screening approaches to ensure that those who are at higher risk of breast cancer receive more intensive screening or screening with different modalities. The continued progress in modalities like tomosynthesis, contrast enhanced mammography and MRI, has the potential to make the transition towards individualized screening more achievable.

In study I, we determine a range of screening mammography benchmarks, that can be useful for comparing the performance of standalone AI systems and selecting an appropriate operating point for AI.

In study II, we evaluate three commercial AI algorithms and conclude that the best performing algorithm could assess screening mammograms with a diagnostic performance exceeding that of radiologists.

In study III, we determine several differences in false assessments between AI and radiologists in mammography screening and highlight that AI can have an important complementary role when combined with radiologists, not only to reduce the risk of ageism amongst women attending screening but also to increase sensitivity for high-density women.

In study IV, we evaluate the efficiency of AI in identifying women with undetected cancer after negative screening mammography. We conclude that using AI to triage women with high AISmartDensity scores to supplemental MRI, diagnoses a considerable number of undetected cancers in mammography screening. Employing AI as a selection method for supplemental MRI screening results in a greater cancer detection rate than using breast density.

To summarize, the results presented in this thesis demonstrate the significant potential of AI in tumor detection. As research and development in this field continues to advance, we can expect further improvements in the accuracy of AI systems, ultimately leading to better patient outcomes and more individualized screening.

8 Ethical considerations

The possibilities for AI within breast radiology are limitless and it is crucial that the ethical considerations are prioritized in the design and implementation of the AI systems.

One of the most significant ethical considerations when implementing AI systems is systemic bias. AI systems are only as good as the data they are trained on, and if that data is biased in any way, the AI system will reflect that bias. For example, if an AI system is trained on data that reflects bias, the AI system will perpetuate those biases, creating inequality in assessments - biased input leading to biased output. An algorithm trained mainly on one ethnic population might not work as well in another ethnic population causing prejudice. To combat this issue, its crucial to ensure that the data used to train AI systems is diverse and representative of all groups without the potential of bias.

Another important ethical consideration when implementing AI systems is transparency. AI systems often operate in opaque ways, which can make it difficult to understand why they make certain decisions. This lack of transparency can lead to distrust and suspicion of AI systems. For the algorithm to be as correct as possible in assessment making one must understand what the algorithm does and why, and also how it was trained and on what data. It's essential to ensure that AI systems are transparent and explainable, to do this we must tackle the "black-box" problem and know why and how AI reached its conclusion.

The question of responsibility has been up for debate many times. Who bears the legal responsibility when a cancer is missed? It all depends on how the algorithm is implemented. If the algorithm only acts as an advisory, the doctor that makes the decision might be the one that bears legal responsibility. However, implementation of tools that streamline assessments may lead to a superstition towards the AI algorithm and the doctor might make decisions based on the algorithm without any real influence; for example when the production requirements are high and the assessments can't be as rigorous. It's crucial to ensure that there is a clear chain of responsibility and accountability for all different scenarios when implementing AI systems in medical imaging.

In conclusion, I believe that the implementation of artificial intelligence can transform breast radiology. However, the ethical discussion must be lifted forward and proper principals must be endorsed before the algorithms are installed in clinical practice.

All studies included in the thesis have an ethical approval from the Ethical review board.

Sub studies I, II and III; DNR: 2016/2600-31

In these two retrospective studies we analyze data from all patients that underwent mammograms in Stockholm county during the years 2008-2015. All data is anonymous. The studies consist of a large number of patients with coded data, securely stored at Karolinska University Hospital Solna, to ensure the safety of patient personal integrity.

Sub study IV; DNR: 2020-00487

It is a randomized clinical trial. Patients undergoing supplemental MRI have given their written consent. All data is securely stored at Karolinska University Hospital Solna, to ensure the safety of patient personal integrity.

9 Points of perspective

Breast cancer is the most common type of cancer affecting women. The last decade there have been many advances concerning treatment, prevention and diagnostics. The progress within breast cancer care will continue in the future and the implementation of artificial intelligence will have a major role. Artificial intelligence has the potential to revolutionize breast radiology by enabling more accurate and efficient diagnosis of breast abnormalities. The use of AI in clinical practice will enhance patientcare, reduce expenses, provide improved decision-making and reduce radiologist workload.

Currently mammography is the only tool used in breast cancer screening. However the last years have brought on an intensive discussion about how breast cancer screening can be made more efficient. I believe that this can be achieved by individualizing the screening program and incorporating AI in the clinical workflow. The latest guidelines from EUSOBI have recommended a shift towards more individualized screening and advocate supplemental screening with MRI for women with dense breasts. Using breast density as an indicator for supplemental MRI screening is promising but might be difficult to implement due to limited MRI resources and increased costs. However, utilizing an artificial intelligence model to further narrow down which women that require additional imaging could be a more sustainable and cost-effective solution.

The last years there has been an uprise in the amount of research regarding artificial intelligence and breast imaging, the study results have shown that AI is a promising tool that have evolved and perform just as well as radiologists and sometimes even better. However, most studies have been retrospective and on limited data. There is a need for more prospective studies where we evaluate AI in a clinical setting to get a better understanding of different problems that might arise and how the implementation will affect the daily workflow.

In conclusion, I believe that the future of AI in breast radiology is bright as it has the potential to improve mammography screening by reducing the inter-reader variation in radiologist assessments and help identify cancers at an earlier stage. Additionally, AI can provide a more accurate estimation of breast cancer risk. While being optimistic, one must also keep in mind that the introduction of AI in breast radiology will come with its challenges. However, with proper infrastructure, maintenance and monitoring we can ensure patient safety and quality patient care.

10 Sammanfattning på Svenska (Swedish summary)

Bröstcancer är den vanligaste formen av cancer bland kvinnor i världen. I Sverige diagnostiseras cirka 9,000 kvinnor med bröstcancer varje år. Dödligheten började minska på 1980- och 1990-talet, då många länder införde mammografiscreening. Mammografiscreening används i flera länder runt om i världen och implementeringen har bidragit till tidigare diagnos vilket har minskat dödligheten med 30-40%. De flesta kvinnor som får en bröstcancer diagnos är över 50 år, men även yngre kvinnor kan drabbas. I de allra flesta fall finns inga identifierbara riskfaktorer förutom ålder och kön. Ytterligare riskfaktorer som är kopplade till bröstcancer är brösttäthet, ärftlighet, reproduktiv hälsa, amning, fetma, hormonterapi och alkohol konsumtion.

Nästan 65% av all bröstcancer upptäcks genom mammografiscreening. I Sverige bjuds alla kvinnor mellan 40-74 år in till mammografiscreening var 18-24 månad. Screeningsprocessen innebär att man tar två bilder av varje bröst och frågar om kliniska symptom från brösten såsom en nytillkommen knöl eller sekretion från bröstvårtan. Alla screeningundersökningar granskas av två bröstradiologer. Om mammografi-undersökningen flaggas på grund av en misstänkt förändring i bröstet eller om kvinnan har kliniska symptom, granskas undersökningen ånyo vid en konsensusdiskussion. Under konsensusdiskussionen granskar minst två erfarna bröstradiologer bilderna och tar beslut om kvinnan är frisk eller om hon behöver kallas tillbaka för vidareutredning.

Mammografiscreeningen står inför många utmaningar, såsom brist på bröstradiologer, betydande variation i bedömningarna mellan radiologer, en lägre sensitivitet för kvinnor med täta bröst samt en hög andel intervallcancrar. Intervallcancer är bröstcancer som upptäcks kliniskt mellan två screeningsundersökningar. De är vanligtvis större vid tidpunkten för upptäckt och de är kopplade till en högre mortalitet. För att lösa dessa utmaningar, kan artificiell intelligens vara ett användbart verktyg. Implementeringen av AI skulle kunna effektivisera screeningsprocessen genom att upptäcka cancer i ett tidigare skede, minska frekvensen av intervallcancer och skräddarsy screeningsprocessen och göra den mer individualiserad.

I denna avhandling har vi analyserat hur man kan använda artificiell intelligens för och ta itu med de ovannämnda utmaningarna.

I studie I, analyserade vi cirka 1,000,000 mammografiscreenings bedömningar från radiologer i Stockholms län. Vi undersökte den övergripande prestandan men även uppdelat efter olika tumörkarakteristika. Resultaten visade på en stor variation i prestanda och sensitiviteten varierade beroende på tumör typ.

I studie II, utvärderade vi prestandan hos tre kommersiella algoritmer och jämförde dem med de retrospektiva bedömningarna från radiologerna i studie I. Resultaten visade att

den algoritm som presterade bäst överträffade radiologerna i bedömningen av screeningundersökningarna.

I studie III, undersökte vi om det fanns oenigheter i bedömningarna mellan AI och radiologerna. Vi undersökte vidare huruvida brösttäthet och tumörkarakteristika påverkade bedömningarna. Resultaten visade att AI kan ha en viktig kompletterande roll när den kombineras med radiologerna, särskilt för kvinnor med hög brösttäthet.

I studie IV, genomför vi en randomiserad klinisk studie för och utvärdera effekten av ett AI-baserat urval till kompletterande screening med MR. Studien pågår fortfarande, men hittills är resultaten hoppgivande. De preliminära resultaten visar att cancerdetektions frekvensen är betydligt högre vid AI-baserat urval än vad som rapporterats för urvalsmetoder baserade på brösttäthet.

Sammanfattningsvis har vi visat att användningen av AI för detektion av bröstcancer kan öka precisionen och effektiviteten inom mammografi-screening.

11 Acknowledgements

Fredrik Strand, my main supervisor. For always believing in me and pushing me forward. For always taking your time and being available no matter how busy you are. For challenging me to do better and become better. For being patient and encouraging when I have wanted to give up. For always being kind and humorous. Without you, this would not have been possible. Thank you for being a great supervisor and a true inspiration!

Martin Eklund, my co-supervisor. For always being so humble and kind. For all the guidance with biostatistics throughout my studies. For all the encouragement and feedback.

Kevin Smith, my co-supervisor. For all the feedback and help with technical parts of the studies. For all the laughter and good times.

Theodoros Foukakis, my co-supervisor. For all the feedback and guidance through the world of breast oncology.

Antonios Valachis, Chikako Suzuki and Emilio Ugalde Morales, the half time committee for your constructive feedback.

Lars Holmgren, Andreas Lundqvist, Erika Rindsjö and the rest of the Onk-Pat department. For your guidance and support throughout this journey.

Karin Dembrower, senior breast radiologist and my mentor throughout this journey. For always taking your time to guide and support me, for always offering a listening ear and a kind heart. For always being so understanding.

Athanasios Zouzos, senior breast radiologist and head of the breast radiology department at Karolinska University Hospital. For your immense patience and support during my research and clinical career. For teaching me, believing in me and always being positive.

Jonaz Ripsweden, senior radiologist and chief of the radiology residency program at Karolinska University Hospital. For all your support, for allowing me to go to research school, conferences and courses. For giving me the time I need to conduct my research and for always believing in me and cheering me on.

Mari-Louise Wretling, Erik Jensen, Koshiar Medson and Elin Boija, the chiefs and directors of the residency program at the Karolinska University Hospital. For all your hard work and dedication into making the radiology residency program better.

Malin Laurell, senior breast radiologist, for taking me in and seeing my potential. For always being so kind and positive.

Edward Azawedo, senior breast radiologist, pathologist and associate professor. For your guidance and kindness, for giving me a great introduction into the field of breast radiology and for continuing to inspire me.

Dimitra Ntoula, senior breast radiologist. For teaching me, supporting me and making me laugh. For always being a ray of sunshine. For being thoughtful and kind.

Afroditi Komoti, breast/thoracic radiologist. For always being positive and humoristic. For listening and giving me life and work advice.

Hedvig Moll, fellow radiology resident and future breast radiology colleague, for being you, for all the laughter and support during work and lunches.

Shirin Rasoul Olyaei and Csilla Srivastava, radiology nurses at the breast MRI department. For all the hard work you do for the study patients. For all the laughter and encouragement during good and bad times.

Erik Wåhlin, physicist. For all your hard work and patience with data collection.

Yue Liu and Moein Sorkhei, fellow PhD students for always helping out with the AI computer and quickly answering my very early morning slack messages.

Yanlu Wang, for always helping out with technical support and for all the laughter.

Elin Annetorp & Kirsi Salo, for always being positive and for all your hard work on the ScreenTrust MRI study.

All the wonderful colleagues at the breast radiology department, Karolinska University hospital. For always working so hard, for always being so positive and for contributing to our work place being one of the best.

All the brilliant researchers in the Fredrik Strand research team. For being helpful and kind. For conducting important research and contributing to a very important cause. And a special thank you to **Fernando Cossio,** for always being available when I've needed help.

Violetta and Grzegorz, my parents in-law. For always supporting me and believing in me.

Izabela, my husband's godmother. For all the motivation and support. For always being so positive and welcoming.

Paulina and Aesar, my sister in-law and her husband. For all the good times and laughter.

Ali, Zeyneb and Noah, my brother, my sister in-law and my nephew. For all the good times and for making me an aunt.

Sally, my sister. For always cheering me on and making me laugh. Ruuf'us on fire!

Hend, my best friend since childhood. For supporting me and cheering me on. For laughing with me and crying with me. For always being here for me even though you are several miles away.

Saleh and Ehlimana, my brother and my sister in-law and my best friends. For always supporting me and having my back. For always being there for me when I need you the most. For all the good memories and for all the laughter we share. You mean the world to me. Life would not be the same without you.

Mom and Dad. For all the hard work you put into raising me, for all the sacrifices you made for me and for giving me the best opportunities in life to prosper. Mom, you are my biggest motivation in life and even though you are no longer with us, your strength is what keeps me going each and every day.

Alexander, my husband. For all the love, support and laughter. For all your patience, for all the times you have pushed me and for all the times you have motivated me. Together we can do it all, I love you!

Nellie, my beautiful daughter. For showing me the true meaning of life, I love you!

12 References

1. Statistics on Cancer Incidence 2019: The National Board of Health and Welfare; 2020 [Internet].

2. WHO. 2020.

3. Folkhälsomyndigheten. 2020.

4. Tabár L, Dean PB, Chen TH, Yen AM, Chen SL, Fann JC, et al. The incidence of fatal breast cancer measures the increased effectiveness of therapy in women participating in mammography screening. Cancer. 2019;125(4):515-23.

5. Duffy SW, Tabár L, Yen AM, Dean PB, Smith RA, Jonsson H, et al. Beneficial Effect of Consecutive Screening Mammography Examinations on Mortality from Breast Cancer: A Prospective Study. Radiology. 2021;299(3):541-7.

6. Seely JM, Alhassan T. Screening for breast cancer in 2018-what should we be doing today? Curr Oncol. 2018;25(Suppl 1):S115-S24.

7. Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M. The benefits and harms of breast cancer screening: an independent review. Br J Cancer. 2013;108(11):2205-40.

8. Tabar L, Vitak B, Chen TH, Yen AM, Cohen A, Tot T, et al. Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. Radiology. 2011;260(3):658-63.

9. Barlow WE, Chi C, Carney PA, Taplin SH, D'Orsi C, Cutter G, et al. Accuracy of Screening Mammography Interpretation by Characteristics of Radiologists. JNCI: Journal of the National Cancer Institute. 2004;96(24):1840-50.

10. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol Biomarkers Prev. 2006;15(6):1159-69.

11. Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, et al. Mammographic density and the risk and detection of breast cancer. N Engl J Med. 2007;356(3):227-36.

12. Benndorf M, Baltzer PAT, Vag T, Gajda M, Runnebaum IB, Kaiser WA. Breast MRI as an adjunct to mammography: Does it really suffer from low specificity? a retrospective analysis stratified by mammographic BI-RADS classes. Acta Radiologica. 2010;51(7):715-21.

13. Hansen JT. Netters clinical anatomy Elsevier health science division; 2018.

14. Atashgaran V, Wrin J, Barry SC, Dasari P, Ingman WV. Dissecting the Biology of Menstrual Cycle-Associated Breast Cancer Risk. Front Oncol. 2016;6:267-.

15. Radisky DC, Hartmann LC. Mammary involution and breast cancer risk: transgenic models and clinical studies. J Mammary Gland Biol Neoplasia. 2009;14(2):181-91.

16. Russo J, Moral R, Balogh GA, Mailo D, Russo IH. The protective role of pregnancy in breast cancer. Breast cancer research : BCR. 2005;7(3):131-42.

17. Sherman BM, Korenman SG. Hormonal characteristics of the human menstrual cycle throughout reproductive life. J Clin Invest. 1975;55(4):699-706.

18. Russo J, Russo IH. Development of the human breast. Maturitas. 2004;49(1):2-15.

19. Gusterson BA, Stein T. Human breast development. Semin Cell Dev Biol. 2012;23(5):567-73.

20. DeSantis CE, Bray F, Ferlay J, Lortet-Tieulent J, Anderson BO, Jemal A. International Variation in Female Breast Cancer Incidence and Mortality Rates. Cancer Epidemiology Biomarkers & Prevention. 2015;24(10):1495-506.

21. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians. 2021;71(3):209-49.

22. RCC. 2020.

23. Cancerfonden. 2020 [Available from: https://www.cancerfonden.se/om-cancer/cancersjukdomar/brostcancer.

24. bröstcancer Nv. 2020 [Available from: https://kunskapsbanken.cancercentrum.se/globalassets/cancerdiagnoser/brost/vardprogram/nationellt-vardprogram-brostcancer.pdf.

25. Division of Cancer Prevention and Control CfDCaP. Breast cancer https://www.cdc.gov/cancer/breast/index.htm2020 [

26. Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, et al. Mammographic Density and the Risk and Detection of Breast Cancer. New England Journal of Medicine. 2007;356(3):227-36.

27. McPherson K, Steel CM, Dixon JM. ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. BMJ. 2000;321(7261):624-8.

28. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. Lancet. 2001;358(9291):1389-99.

29. Colditz GA, Kaphingst KA, Hankinson SE, Rosner B. Family history and risk of breast cancer: nurses' health study. Breast cancer research and treatment. 2012;133(3):1097-104.

30. Singletary SE. Rating the risk factors for breast cancer. Ann Surg. 2003;237(4):474-82.

31. Brinton LA, Schairer C, Hoover RN, Fraumeni JF, Jr. Menstrual factors and risk of breast cancer. Cancer Invest. 1988;6(3):245-54.

32. Titus-Ernstoff L, Longnecker MP, Newcomb PA, Dain B, Greenberg ER, Mittendorf R, et al. Menstrual factors in relation to breast cancer risk. Cancer Epidemiol Biomarkers Prev. 1998;7(9):783-9.

33. Shiovitz S, Korde LA. Genetics of breast cancer: a topic in evolution. Ann Oncol. 2015;26(7):1291-9.

34. Krammer J, Pinker-Domenig K, Robson ME, Gönen M, Bernard-Davila B, Morris EA, et al. Breast cancer detection and tumor characteristics in BRCA1 and BRCA2 mutation carriers. Breast Cancer Res Treat. 2017;163(3):565-71.

35. Foulkes WD. Inherited susceptibility to common cancers. N Engl J Med. 2008;359(20):2143-53.

36. Hinkula M, Pukkala E, Kyyrönen P, Kauppila A. Grand multiparity and the risk of breast cancer: population-based study in Finland. Cancer Causes Control. 2001;12(6):491-500.

37. Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50302 women with breast cancer and 96973 women without the disease. Lancet. 2002;360(9328):187-95.

38. D'Orsi CJ SE, Mendelson EB et al. ACR BI-RADS ® Atlas, Breast Imaging Reporting and Data System. American College of Radiology, Reston, VA. 2013.

39. Eberl MM, Fox CH, Edge SB, Carter CA, Mahoney MC. BI-RADS classification for management of abnormal mammograms. J Am Board Fam Med. 2006;19(2):161-4.

40. Freer PE. Mammographic Breast Density: Impact on Breast Cancer Risk and Implications for Screening. RadioGraphics. 2015;35(2):302-15.

41. Price ER, Hargreaves J, Lipson JA, Sickles EA, Brenner RJ, Lindfors KK, et al. The California Breast Density Information Group: A Collaborative Response to the Issues of Breast Density, Breast Cancer Risk, and Breast Density Notification Legislation. Radiology. 2013;269(3):887-92.

42. Mammographic density: a hormonally responsive risk factor for breast cancer. British Menopause Society Journal. 2006;12(4):186-93.

43. Engmann NJ, Golmakani MK, Miglioretti DL, Sprague BL, Kerlikowske K, Consortium ftBCS. Population-Attributable Risk Proportion of Clinical Risk Factors for Breast Cancer. JAMA Oncology. 2017;3(9):1228-36.

44. Brentnall AR, Cuzick J, Buist DSM, Bowles EJA. Long-term Accuracy of Breast Cancer Risk Assessment Combining Classic Risk Factors and Breast Density. JAMA Oncology. 2018;4(9):e180174-e.

45. Buist DSM, Porter PL, Lehman C, Taplin SH, White E. Factors Contributing to Mammography Failure in Women Aged 40–49 Years. JNCI: Journal of the National Cancer Institute. 2004;96(19):1432-40.

46. Saadatmand S, Bretveld R, Siesling S, Tilanus-Linthorst MM. Influence of tumour stage at breast cancer detection on survival in modern times: population based study in 173,797 patients. Bmj. 2015;351:h4901.

47. O'Sullivan B, Brierley J, Byrd D, Bosman F, Kehoe S, Kossary C, et al. The TNM classification of malignant tumours-towards common understanding and reasonable expectations. Lancet Oncol. 2017;18(7):849-51.

48. Rosen RD SATC. 2022.

49. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology. 1991;19(5):403-10.

50. Leonard GD, Swain SM. Ductal carcinoma in situ, complexities and challenges. J Natl Cancer Inst. 2004;96(12):906-20.

51. Espina V, Liotta LA. What is the malignant nature of human ductal carcinoma in situ? Nat Rev Cancer. 2011;11(1):68-75.

52. Cowell CF, Weigelt B, Sakr RA, Ng CK, Hicks J, King TA, et al. Progression from ductal carcinoma in situ to invasive breast cancer: revisited. Mol Oncol. 2013;7(5):859-69.

53. Perry N, Broeders M, de Wolf C, Törnberg S, Holland R, von Karsa L. European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth edition--summary document. Ann Oncol. 2008;19(4):614-22.

54. Barnes NL, Dimopoulos N, Williams KE, Howe M, Bundred NJ. The frequency of presentation and clinico-pathological characteristics of symptomatic versus screen detected ductal carcinoma in situ of the breast. Eur J Surg Oncol. 2014;40(3):249-54.

55. Makki J. Diversity of Breast Carcinoma: Histological Subtypes and Clinical Relevance. Clin Med Insights Pathol. 2015;8:23-31.

56. Badowska-Kozakiewicz AM, Liszcz A, Sobol M, Patera J. Retrospective evaluation of histopathological examinations in invasive ductal breast cancer of no special type: an analysis of 691 patients. Arch Med Sci. 2017;13(6):1408-15.

57. Abe O, Abe R, Enomoto K, Kikuchi K, Koyama H, Masuda H, et al. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. 2005.

58. Bentzon N, Düring M, Rasmussen BB, Mouridsen H, Kroman N. Prognostic effect of estrogen receptor status across age in primary breast cancer. Int J Cancer. 2008;122(5):1089-94.

59. Barnes DM, Hanby AM. Oestrogen and progesterone receptors in breast cancer: past, present and future. Histopathology. 2001;38(3):271-4.

60. Hawkins RA, Roberts MM, Forrest AP. Oestrogen receptors and breast cancer: current status. Br J Surg. 1980;67(3):153-69.

61. Iqbal N, Iqbal N. Human Epidermal Growth Factor Receptor 2 (HER2) in Cancers: Overexpression and Therapeutic Implications. Mol Biol Int. 2014;2014:852748.

62. Soliman NA, Yussif SM. Ki-67 as a prognostic marker according to breast cancer molecular subtype. Cancer Biol Med. 2016;13(4):496-504.

63. de Azambuja E, Cardoso F, de Castro G, Jr., Colozza M, Mano MS, Durbecq V, et al. Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12,155 patients. Br J Cancer. 2007;96(10):1504-13.

64. Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thurlimann B, et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Ann Oncol. 2013;24(9):2206-23.

65. Hon JD, Singh B, Sahin A, Du G, Wang J, Wang VY, et al. Breast cancer molecular subtypes: from TNBC to QNBC. Am J Cancer Res. 2016;6(9):1864-72.

66. Moo TA, Sanford R, Dang C, Morrow M. Overview of Breast Cancer Therapy. PET Clin. 2018;13(3):339-54.

67. Litière S, Werutsky G, Fentiman IS, Rutgers E, Christiaens MR, Van Limbergen E, et al. Breast conserving therapy versus mastectomy for stage I-II breast cancer:

20 year follow-up of the EORTC 10801 phase 3 randomised trial. Lancet Oncol. 2012;13(4):412-9.

68. Moran MS, Schnitt SJ, Giuliano AE, Harris JR, Khan SA, Horton J, et al. Society of Surgical Oncology-American Society for Radiation Oncology consensus guideline on margins for breast-conserving surgery with whole-breast irradiation in stages I and II invasive breast cancer. Int J Radiat Oncol Biol Phys. 2014;88(3):553-64.

69. Tamburelli F, Maggiorotto F, Marchiò C, Balmativola D, Magistris A, Kubatzki F, et al. Reoperation rate after breast conserving surgery as quality indicator in breast cancer treatment: A reappraisal. Breast. 2020;53:181-8.

70. Fisher B, Anderson S, Bryant J, Margolese RG, Deutsch M, Fisher ER, et al. Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. N Engl J Med. 2002;347(16):1233-41.

71. Darby S, McGale P, Correa C, Taylor C, Arriagada R, Clarke M, et al. Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10,801 women in 17 randomised trials. Lancet. 2011;378(9804):1707-16.

72. Burstein HJ, Curigliano G, Loibl S, Dubsky P, Gnant M, Poortmans P, et al. Estimating the benefits of therapy for early-stage breast cancer: the St. Gallen International Consensus Guidelines for the primary therapy of early breast cancer 2019. Ann Oncol. 2019;30(10):1541-57.

73. Li J, Shao Z. Mammography screening in less developed countries. Springerplus. 2015;4:615.

74. Duffy SW, Tabár L, Yen AM-F, Dean PB, Smith RA, Jonsson H, et al. Mammography screening reduces rates of advanced and fatal breast cancers: Results in 549,091 women. Cancer. 2020;126(13):2971-9.

75. Bergh J, Jönsson PE, Glimelius B, Nygren P. A systematic overview of chemotherapy effects in breast cancer. Acta Oncol. 2001;40(2-3):253-81.

76. Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med. 2005;353(17):1784-92.

77. Zheng Y-Z, Wang L, Hu X, Shao Z-M. Effect of tumor size on breast cancer-specific survival stratified by joint hormone receptor status in a SEER population-based study. Oncotarget. 2015;6(26):22985-95.

78. Fitzgibbons PL, Page DL, Weaver D, Thor AD, Allred DC, Clark GM, et al. Prognostic factors in breast cancer. College of American Pathologists Consensus Statement 1999. Arch Pathol Lab Med. 2000;124(7):966-78.

79. Mirza AN, Mirza NQ, Vlastos G, Singletary SE. Prognostic factors in node-negative breast cancer: a review of studies with sample size more than 200 and follow-up more than 5 years. Ann Surg. 2002;235(1):10-26.

80. Rosenberg J, Chia YL, Plevritis S. The effect of age, race, tumor size, tumor grade, and disease stage on invasive ductal breast cancer survival in the U.S. SEER database. Breast Cancer Res Treat. 2005;89(1):47-54.

81. Untch M, Thomssen C, Bauerfeind I, Braun M, Brucker SY, Felberbaum R, et al. Primary Therapy of Early Breast Cancer: Evidence, Controversies, Consensus: Spectrum of Opinion of German Specialists on the 16th St. Gallen International Breast Cancer Conference (Vienna 2019). Geburtshilfe Frauenheilkd. 2019;79(6):591-604.

82. Yaffe MJ. AAPM tutorial. Physics of mammography: image recording process. RadioGraphics. 1990;10(2):341-63.

83. Tabár L, Gad A, Holmberg LH, Ljungquist U, Fagerberg CJG, Baldetorp L, et al. REDUCTION IN MORTALITY FROM BREAST CANCER AFTER MASS SCREENING WITH MAMMOGRAPHY: Randomised Trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. The Lancet. 1985;325(8433):829-32.

84. Lynge E, Vejborg I, Andersen Z, von Euler-Chelpin M, Napolitano G. Mammographic Density and Screening Sensitivity, Breast Cancer Incidence and Associated Risk Factors in Danish Breast Cancer Screening. J Clin Med. 2019;8(11):2021.

85. Gierach GL, Ichikawa L, Kerlikowske K, Brinton LA, Farhat GN, Vacek PM, et al. Relationship between mammographic density and breast cancer death in the Breast Cancer Surveillance Consortium. Journal of the National Cancer Institute. 2012;104(16):1218-27.

86. Aiello EJ, Buist DS, White E, Porter PL. Association between mammographic breast density and breast cancer tumor characteristics. Cancer Epidemiol Biomarkers Prev. 2005;14(3):662-8.

87. Salim M, Dembrower K, Eklund M, Lindholm P, Strand F. Range of Radiologist Performance in a Population-based Screening Cohort of 1 Million Digital Mammography Examinations. Radiology. 2020;297(1):33-9.

88. Johnson K, Sarma D, Hwang ES. Lobular breast cancer series: imaging. Breast Cancer Research. 2015;17(1):94.

89. Porter AJ, Evans EB, Foxcroft LM, Simpson PT, Lakhani SR. Mammographic and ultrasound features of invasive lobular carcinoma of the breast. Journal of Medical Imaging and Radiation Oncology. 2014;58(1):1-10.

90. Pisano ED, Hendrick RE, Yaffe MJ, Baum JK, Acharyya S, Cormack JB, et al. Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in DMIST. Radiology. 2008;246(2):376-83.

91. Thurfjell MG, Vitak B, Azavedo E, Svane G, Thurfjell E. Effect on sensitivity and specificity of mammography screening with or without comparison of old mammograms. Acta Radiol. 2000;41(1):52-6.

92. Kemp Jacobsen K, O'Meara ES, Key D, D SMB, Kerlikowske K, Vejborg I, et al. Comparing sensitivity and specificity of screening mammography in the United States and Denmark. Int J Cancer. 2015;137(9):2198-207.

93. Baines CJ, Miller AB, Wall C, McFarlane DV, Simor IS, Jong R, et al. Sensitivity and specificity of first screen mammography in the Canadian National Breast Screening Study: a preliminary report from five centers. Radiology. 1986;160(2):295-8.

94. Ciatto S, Visioli C, Paci E, Zappa M. Breast density as a determinant of interval cancer at mammographic screening. Br J Cancer. 2004;90(2):393-6.

95. Weigel S, Heindel W, Heidrich J, Hense HW, Heidinger O. Digital mammography screening: sensitivity of the programme dependent on breast density. Eur Radiol. 2017;27(7):2744-51.

96. Törnberg S, Kemetli L, Ascunce N, Hofvind S, Anttila A, Sèradour B, et al. A pooled analysis of interval cancer rates in six European countries. European Journal of Cancer Prevention. 2010;19(2):87-93.

97. Holm J, Humphreys K, Li J, Ploner A, Cheddad A, Eriksson M, et al. Risk Factors and Tumor Characteristics of Interval Cancers by Mammographic Density. Journal of Clinical Oncology. 2015;33(9):1030-7.

98. Strand F, Humphreys K, Holm J, Eriksson M, Törnberg S, Hall P, et al. Long-term prognostic implications of risk factors associated with tumor size: a case study of women regularly attending screening. Breast cancer research : BCR. 2018;20(1):31-.

99. Altobelli E, Rapacchietta L, Angeletti PM, Barbante L, Profeta FV, Fagnano R. Breast Cancer Screening Programmes across the WHO European Region: Differences among Countries Based on National Income Level. Int J Environ Res Public Health. 2017;14(4).

100. Liberman L, Menell JH. Breast imaging reporting and data system (BI-RADS). Radiol Clin North Am. 2002;40(3):409-30, v.

101. mammogram ECIoBCFaat. 2021.

102. Li T, Marinovich ML, Houssami N. Digital breast tomosynthesis (3D mammography) for breast cancer screening and for assessment of screen-recalled findings: review of the evidence. Expert Rev Anticancer Ther. 2018;18(8):785-91.

103. Zackrisson S, Lång K, Rosso A, Johnson K, Dustler M, Förnvik D, et al. One-view breast tomosynthesis versus two-view mammography in the Malmö Breast Tomosynthesis Screening Trial (MBTST): a prospective, population-based, diagnostic accuracy study. Lancet Oncol. 2018;19(11):1493-503.

104. Lång K, Andersson I, Rosso A, Tingberg A, Timberg P, Zackrisson S. Performance of one-view breast tomosynthesis as a stand-alone breast cancer screening modality: results from the Malmö Breast Tomosynthesis Screening Trial, a population-based study. Eur Radiol. 2016;26(1):184-90.

105. Conant EF, Zuckerman SP, McDonald ES, Weinstein SP, Korhonen KE, Birnbaum JA, et al. Five Consecutive Years of Screening with Digital Breast Tomosynthesis: Outcomes by Screening Year and Round. Radiology. 2020;295(2):285-93.

106. Conant EF, Barlow WE, Herschorn SD, Weaver DL, Beaber EF, Tosteson ANA, et al. Association of Digital Breast Tomosynthesis vs Digital Mammography With Cancer Detection and Recall Rates by Age and Breast Density. JAMA Oncol. 2019;5(5):635-42.

107. Conant EF, Talley MM, Parghi CR, Sheh BC, Liang S-Y, Pohlman S, et al. Mammographic Screening in Routine Practice: Multisite Study of Digital Breast Tomosynthesis and Digital Mammography Screenings. Radiology.0(0):221571.

108. Cozzi A, Magni V, Zanardo M, Schiaffino S, Sardanelli F. Contrast-enhanced Mammography: A Systematic Review and Meta-Analysis of Diagnostic Performance. Radiology. 2022;302(3):568-81.

109. Sorin V, Yagil Y, Yosepovich A, Shalmon A, Gotlieb M, Neiman OH, et al. Contrast-Enhanced Spectral Mammography in Women With Intermediate Breast Cancer Risk and Dense Breasts. AJR Am J Roentgenol. 2018;211(5):W267-w74.

110. Jochelson MS, Lobbes MBI. Contrast-enhanced Mammography: State of the Art. Radiology. 2021;299(1):36-48.

111. Pötsch N, Vatteroni G, Clauser P, Helbich TH, Baltzer PAT. Contrast-enhanced Mammography versus Contrast-enhanced Breast MRI: A Systematic Review and Meta-Analysis. Radiology. 2022;305(1):94-103.

112. Teh W, Wilson AR. The role of ultrasound in breast cancer screening. A consensus statement by the European Group for Breast Cancer Screening. Eur J Cancer. 1998;34(4):449-50.

113. Rebolj M, Assi V, Brentnall A, Parmar D, Duffy SW. Addition of ultrasound to mammography in the case of dense breast tissue: systematic review and meta-analysis. Br J Cancer. 2018;118(12):1559-70.

114. Sprague BL, Stout NK, Schechter C, van Ravesteyn NT, Cevik M, Alagoz O, et al. Benefits, harms, and cost-effectiveness of supplemental ultrasonography screening for women with dense breasts. Ann Intern Med. 2015;162(3):157-66.

115. Kolb TM, Lichy J, Newhouse JH. Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations. Radiology. 2002;225(1):165-75.

116. Buchberger W, Niehoff A, Obrist P, DeKoekkoek-Doll P, Dünser M. Clinically and mammographically occult breast lesions: detection and classification with high-resolution sonography. Semin Ultrasound CT MR. 2000;21(4):325-36.

117. Crystal P, Strano SD, Shcharynski S, Koretz MJ. Using sonography to screen women with mammographically dense breasts. AJR Am J Roentgenol. 2003;181(1):177-82.

118. Gordon PB, Goldenberg SL. Malignant breast masses detected only by ultrasound. A retrospective review. Cancer. 1995;76(4):626-30.

119. Kaplan SS. Clinical utility of bilateral whole-breast US in the evaluation of women with dense breast tissue. Radiology. 2001;221(3):641-9.

120. Leconte I, Feger C, Galant C, Berlière M, Berg BV, D'Hoore W, et al. Mammography and subsequent whole-breast sonography of nonpalpable breast cancers: the importance of radiologic breast density. AJR Am J Roentgenol. 2003;180(6):1675-9.

121. Corsetti V, Ferrari A, Ghirardi M, Bergonzini R, Bellarosa S, Angelini O, et al. Role of ultrasonography in detecting mammographically occult breast carcinoma in women with dense breasts. Radiol Med. 2006;111(3):440-8.

122. Berg WA, Blume JD, Cormack JB, Mendelson EB, Lehrer D, Böhm-Vélez M, et al. Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. Jama. 2008;299(18):2151-63.

123. Morrow M, Waters J, Morris E. MRI for breast cancer screening, diagnosis, and treatment. Lancet. 2011;378(9805):1804-11.

124. Mann RM, Kuhl CK, Moy L. Contrast-enhanced MRI for breast cancer screening. J Magn Reson Imaging. 2019;50(2):377-90.

125. Orel SG, Schnall MD. MR imaging of the breast for the detection, diagnosis, and staging of breast cancer. Radiology. 2001;220(1):13-30.

126. Drew PJ, Chatterjee S, Turnbull LW, Read J, Carleton PJ, Fox JN, et al. Dynamic contrast enhanced magnetic resonance imaging of the breast is superior to triple assessment for the pre-operative detection of multifocal breast cancer. Ann Surg Oncol. 1999;6(6):599-603.

127. Fischer U, Kopka L, Grabbe E. Breast carcinoma: effect of preoperative contrast-enhanced MR imaging on the therapeutic approach. Radiology. 1999;213(3):881-8.

128. Heywang-Köbrunner SH, Bick U, Bradley WG, Jr., Boné B, Casselman J, Coulthard A, et al. International investigation of breast MRI: results of a multicentre study (11 sites) concerning diagnostic parameters for contrast-enhanced MRI based on 519 histopathologically correlated lesions. Eur Radiol. 2001;11(4):531-46.

129. Rieber A, Merkle E, Böhm W, Brambs HJ, Tomczak R. MRI of histologically confirmed mammary carcinoma: clinical relevance of diagnostic procedures for detection of multifocal or contralateral secondary carcinoma. J Comput Assist Tomogr. 1997;21(5):773-9.

130. Mann RM, Hoogeveen YL, Blickman JG, Boetes C. MRI compared to conventional diagnostic work-up in the detection and evaluation of invasive lobular carcinoma of the breast: a review of existing literature. Breast Cancer Res Treat. 2008;107(1):1-14.

131. Berg WA, Zhang Z, Lehrer D, Jong RA, Pisano ED, Barr RG, et al. Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk. Jama. 2012;307(13):1394-404.

132. Veenhuizen SGA, de Lange SV, Bakker MF, Pijnappel RM, Mann RM, Monninkhof EM, et al. Supplemental Breast MRI for Women with Extremely Dense Breasts: Results of the Second Screening Round of the DENSE Trial. Radiology. 2021;299(2):278-86.

133. Hussein H, Abbas E, Keshavarzi S, Fazelzad R, Bukhanov K, Kulkarni S, et al. Supplemental Breast Cancer Screening in Women with Dense Breasts and Negative Mammography: A Systematic Review and Meta-Analysis. Radiology.0(0):221785.

134. Moore SG, Shenoy PJ, Fanucchi L, Tumeh JW, Flowers CR. Cost-effectiveness of MRI compared to mammography for breast cancer screening in a high risk population. BMC health services research. 2009;9:9-.

135. Ko ES, Morris EA. Abbreviated Magnetic Resonance Imaging for Breast Cancer Screening: Concept, Early Results, and Considerations. Korean J Radiol. 2019;20(4):533-41.

136. Partovi S, Sin D, Lu Z, Sieck L, Marshall H, Pham R, et al. Fast MRI breast cancer screening - Ready for prime time. Clin Imaging. 2020;60(2):160-8.

137. Baxter GC, Selamoglu A, Mackay JW, Bond S, Gray E, Gilbert FJ. A meta-analysis comparing the diagnostic performance of abbreviated MRI and a full diagnostic protocol in breast cancer. Clinical Radiology. 2021;76(2):154. e23-. e32.

138. Mazari FAK, Sharma N, Reid D, Horgan K. The need for triple assessment and predictors for diagnosis of breast cancer in patients <40 years of age. Clin Radiol. 2018;73(8):758.e19-.e25.

139. Morris A, Pommier RF, Schmidt WA, Shih RL, Alexander PW, Vetto JT. Accurate evaluation of palpable breast masses by the triple test score. Arch Surg. 1998;133(9):930-4.

140. Evidence relevant to guidelines for the investigation of breast symptoms. Second edition, February 2006. Camperdown, NSW: National Breast Cancer Centre; 2006. https://www.canceraustralia.gov.au/sites/default/files/publications/ibsr-evidence-relevant-to-guidelines-for-the-investigation-of-breast-symptoms_504af0324b24a.pdf2006 [

141. Cao C, Liu F, Tan H, Song D, Shu W, Li W, et al. Deep Learning and Its Applications in Biomedicine. Genomics Proteomics Bioinformatics. 2018;16(1):17-32.

142. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436-44.

143. Krizhevsky A, Sutskever I, Hinton G. B Advances in Neural Information Processing Systems 25 (Proc. NIPS 2012),(USA, NV, Stateline, 3-8 December, 2012), USA, NY, Red Hook, Publ. 2012.

144. Farabet C, Couprie C, Najman L, LeCun Y. Learning hierarchical features for scene labeling. IEEE transactions on pattern analysis and machine intelligence. 2012;35(8):1915-29.

145. Hinton G, Deng L, Yu D, Dahl GE, Mohamed A-r, Jaitly N, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal processing magazine. 2012;29(6):82-97.

146. Akselrod-Ballin A, Chorev M, Shoshan Y, Spiro A, Hazan A, Melamed R, et al. Predicting Breast Cancer by Applying Deep Learning to Linked Health Records and Mammograms. Radiology. 2019;292(2):331-42.

147. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, Broeders M, Gennaro G, Clauser P, et al. Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists. JNCI: Journal of the National Cancer Institute. 2019.

148. Lehman CD, Wellman RD, Buist DSM, Kerlikowske K, Tosteson ANA, Miglioretti DL, et al. Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection. JAMA Internal Medicine. 2015;175(11):1828-37.

149. Lehman CD, Arao RF, Sprague BL, Lee JM, Buist DS, Kerlikowske K, et al. National Performance Benchmarks for Modern Screening Digital Mammography: Update from the Breast Cancer Surveillance Consortium. Radiology. 2017;283(1):49-58.

150. Fenton JJ, Taplin SH, Carney PA, Abraham L, Sickles EA, D'Orsi C, et al. Influence of Computer-Aided Detection on Performance of Screening Mammography. New England Journal of Medicine. 2007;356(14):1399-409.

151. Gilbert FJ, Astley SM, Gillan MG, Agbaje OF, Wallis MG, James J, et al. CADET II: A prospective trial of computer-aided detection (CAD) in the UK Breast Screening Programme. Journal of Clinical Oncology. 2008;26(15_suppl):508-.

152. Azavedo E, Zackrisson S, Mejàre I, Heibert Arnlind M. Is single reading with computer-aided detection (CAD) as good as double reading in mammography screening? A systematic review. BMC Med Imaging. 2012;12:22.

153. Ciatto S, Del Turco MR, Risso G, Catarzi S, Bonardi R, Viterbo V, et al. Comparison of standard reading and computer aided detection (CAD) on a national

proficiency test of screening mammography. European Journal of Radiology. 2003;45(2):135-8.

154. O'Donoghue C, Eklund M, Ozanne EM, Esserman LJ. Aggregate cost of mammography screening in the United States: comparison of current practice and advocated guidelines. Ann Intern Med. 2014;160(3):145.

155. Kyono T, Gilbert FJ, van der Schaar M. Improving Workflow Efficiency for Mammography Using Machine Learning. J Am Coll Radiol. 2020;17(1 Pt A):56-63.

156. Ribli D, Horváth A, Unger Z, Pollner P, Csabai I. Detecting and classifying lesions in mammograms with Deep Learning. Sci Rep. 2018;8(1):4165.

157. Salim M, Wahlin E, Dembrower K, Azavedo E, Foukakis T, Liu Y, et al. External Evaluation of 3 Commercial Artificial Intelligence Algorithms for Independent Assessment of Screening Mammograms. JAMA Oncol. 2020;6(10):1581-8.

158. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. Nature. 2020;577(7788):89-94.

159. Leibig C, Brehmer M, Bunk S, Byng D, Pinker K, Umutlu L. Combining the strengths of radiologists and AI for breast cancer screening: a retrospective analysis. The Lancet Digital Health. 2022;4(7):e507-e19.

160. Freeman K, Geppert J, Stinton C, Todkill D, Johnson S, Clarke A, et al. Use of artificial intelligence for image analysis in breast cancer screening programmes: systematic review of test accuracy. BMJ. 2021;374:n1872.

161. Cummings SR, Tice JA, Bauer S, Browner WS, Cuzick J, Ziv E, et al. Prevention of breast cancer in postmenopausal women: approaches to estimating and reducing risk. J Natl Cancer Inst. 2009;101(6):384-98.

162. 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):60-6.

163. Kallenberg M, Petersen K, Nielsen M, Ng AY, Pengfei D, Igel C, et al. Unsupervised Deep Learning Applied to Breast Density Segmentation and Mammographic Risk Scoring. IEEE Trans Med Imaging. 2016;35(5):1322-31.

164. Ha R, Chang P, Karcich J, Mutasa S, Pascual Van Sant E, Liu MZ, et al. Convolutional Neural Network Based Breast Cancer Risk Stratification Using a Mammographic Dataset. Acad Radiol. 2019;26(4):544-9.

165. Zhu X, Wolfgruber TK, Leong L, Jensen M, Scott C, Winham S, et al. Deep Learning Predicts Interval and Screening-detected Cancer from Screening Mammograms: A Case-Case-Control Study in 6369 Women. Radiology. 2021;301(3):550-8.

166. Lotter W, Diab AR, Haslam B, Kim JG, Grisot G, Wu E, et al. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. Nat Med. 2021;27(2):244-9.

167. Yala A, Mikhael PG, Strand F, Lin G, Satuluru S, Kim T, et al. Multi-Institutional Validation of a Mammography-Based Breast Cancer Risk Model. Journal of Clinical Oncology. 2022;40(16):1732-40.

168. Eriksson M, Czene K, Pawitan Y, Leifland K, Darabi H, Hall P. A clinical model for identifying the short-term risk of breast cancer. Breast Cancer Res. 2017;19(1):29.

169. Tan M, Zheng B, Leader JK, Gur D. Association Between Changes in Mammographic Image Features and Risk for Near-Term Breast Cancer Development. IEEE Trans Med Imaging. 2016;35(7):1719-28.

170. Lehman CD, Yala A, Schuster T, Dontchos B, Bahl M, Swanson K, et al. Mammographic Breast Density Assessment Using Deep Learning: Clinical Implementation. Radiology. 2019;290(1):52-8.

171. Magni V, Interlenghi M, Cozzi A, Alì M, Salvatore C, Azzena AA, et al. Development and Validation of an AI-driven Mammographic Breast Density Classification Tool Based on Radiologist Consensus. Radiology: Artificial Intelligence. 2022;4(2):e210199.

172. Strand F. CSAW-CC (mammography). 2022.

173. 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. 2019:10.1007/s10278-019-00278-0.

174. Mattsson B, Wallgren A. Completeness of the Swedish Cancer Register. Non-notified cancer cases recorded on death certificates in 1978. Acta Radiol Oncol. 1984;23(5):305-13.

175. Beam CA, Conant EF, Sickles EA. Association of volume and volume-independent factors with accuracy in screening mammogram interpretation. J Natl Cancer Inst. 2003;95(4):282-90.

176. Molins E, Macia F, Ferrer F, Maristany MT, Castells X. Association between radiologists' experience and accuracy in interpreting screening mammograms. BMC Health Serv Res. 2008;8:91.

177. Esserman L, Cowley H, Eberle C, Kirkpatrick A, Chang S, Berbaum K, et al. Improving the accuracy of mammography: volume and outcome relationships. J Natl Cancer Inst. 2002;94(5):369-75.

178. Burnside ES, Vulkan D, Blanks RG, Duffy SW. Association between Screening Mammography Recall Rate and Interval Cancers in the UK Breast Cancer Service Screening Program: A Cohort Study. Radiology. 2018;288(1):47-54.

179. Ernster VL, Ballard-Barbash R, Barlow WE, Zheng Y, Weaver DL, Cutter G, et al. Detection of Ductal Carcinoma In Situ in Women Undergoing Screening Mammography. JNCI: Journal of the National Cancer Institute. 2002;94(20):1546-54.

180. Dershaw DD, Abramson A, Kinne DW. Ductal carcinoma in situ: mammographic findings and clinical implications. Radiology. 1989;170(2):411-5.

181. Dogan BE, Gonzalez-Angulo AM, Gilcrease M, Dryden MJ, Yang WT. Multimodality Imaging of Triple Receptor–Negative Tumors With Mammography, Ultrasound, and MRI. American Journal of Roentgenology. 2010;194(4):1160-6.

182. Buist DS, Porter PL, Lehman C, Taplin SH, White E. Factors contributing to mammography failure in women aged 40-49 years. J Natl Cancer Inst. 2004;96(19):1432-40.

183. Klemi PJ, Toikkanen S, Räsänen O, Parvinen I, Joensuu H. Mammography screening interval and the frequency of interval cancers in a population-based screening. British Journal of Cancer. 1997;75(5):762-6.

184. Tabár L, Faberberg G, Day NE, Holmberg L. What is the optimum interval between mammographic screening examinations — An analysis based on the latest results of

the Swedish two-county breast cancer screening trial. British Journal of Cancer. 1987;55(5):547-51.

185. Nelson HD, O'Meara ES, Kerlikowske K, Balch S, Miglioretti D. Factors Associated With Rates of False-Positive and False-Negative Results From Digital Mammography Screening: An Analysis of Registry Data. Ann Intern Med. 2016;164(4):226-35.

186. Ho T-QH, Bissell MCS, Kerlikowske K, Hubbard RA, Sprague BL, Lee CI, et al. Cumulative Probability of False-Positive Results After 10 Years of Screening With Digital Breast Tomosynthesis vs Digital Mammography. JAMA Network Open. 2022;5(3):e222440-e.

187. Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V. Effect of Age, Breast Density, and Family History on the Sensitivity of First Screening Mammography. JAMA. 1996;276(1):33-8.

188. Chen SQ, Huang M, Shen YY, Liu CL, Xu CX. Application of Abbreviated Protocol of Magnetic Resonance Imaging for Breast Cancer Screening in Dense Breast Tissue. Acad Radiol. 2017;24(3):316-20.

189. Bao Y, Kwok C, Lee CF. Breast cancer screening behaviors among Chinese women in Mainland China. Nursing & Health Sciences. 2018;20(4):445-51.

190. Ellis L, Canchola AJ, Spiegel D, Ladabaum U, Haile R, Gomez SL. Racial and ethnic disparities in cancer survival: the contribution of tumor, sociodemographic, institutional, and neighborhood characteristics. Journal of Clinical Oncology. 2018;36(1):25.

191. Fan L, Strasser-Weippl K, Li J-J, St Louis J, Finkelstein DM, Yu K-D, et al. Breast cancer in China. The lancet oncology. 2014;15(7):e279-e89.