LUND UNIVERSITY

Master of Science Thesis
2026

AI Lesion Risk Score with DBT and DM

Jonas Magnusson

Supervisors

Anders Tingberg and Magnus Dustler

Medical Radiation Physics, Lund
Medical Physics Programme
Faculty of Science
Lund University, Sweden
www.msf.lu.se

Populärvetenskaplig sammanfattning

Bröstcancer är den vanligaste cancerdiagnosen bland kvinnor i Sverige. Mellan 2013 och 2023 fick ungefär 7900 kvinnor en bröstcancerdiagnos årligen, ungefär 1400 dog som konsekvens. Därför erbjuds möjlighet att delta i en screeningundersökning. I denna undersökning röntgas brösten med målet att upptäcka eventuella tumörer, dessa röntgenbilder kallas för digitala mammografier (DM). En annan form av bildtagning är digital brösttomosyntes (DBT). Vid DBT-bildtagning rör sig röntgenkällan i en båge över bröstet medan bildtagning sker, då kan 3-dimensionella figurer av bröstet skapas. DBT bildtagning har tidigare visats ha en ökad sensitivitet för att upptäcka bröstcancer. Trots detta införs inte DBT i Sverige för screening, bland annat för att radiologerna måste lägga betydligt mer tid på att granska bilderna.

Ett sätt att minska arbetsbördan för radiologerna som blivit alltmer populär är införandet av artificiell intelligens (AI) för att granska dessa bilder. För att detta ska kunna ske måste robusta kvalitetskontrollmetoder kunna införas. Målet med detta projekt var att undersöka en sådan metod.

I studien har upprepade bilder (både DM och DBT) tagits av ett bröstliknande objekt. Bildtagning gjordes med flera olika mammografisystem, samt flera olika inställningar. Det bröstlikande objektet kunde avbildas både med och utan en struktur som efterliknade en tumör. Efter bildtagningen analyserades bilderna med hjälp av AI, som tilldelade en riskpoäng som speglade sannolikheten för att cancer fanns i bilden.

Studien påvisade en markant spridning av resultaten vid bildtagning med DM, även när alla inställningar var identiska. Denna spridning var för två av tillverkarna betydligt lägre vid bildtagning med DBT. En annan sak som upptäcktes var att det fanns en betydligt högre risk för AI:n att felidentifiera frisk vävnad i objektet som tumörvävnad vid bildtagning med DM än för DBT. Allmänt tenderade AI:n även att sätta högre riskpoäng vid bildtagning med DBT än DM om en cancerliknande struktur fanns i objektet.

Abstract

Purpose: The purpose of this project was to examine how the degree of cancer suspicion evaluated by an AI system differs when examining repeated images of a realistic breast phantom acquired with DM and DBT. This project is a part of a larger project to develop quality assurance systems and methods for AI systems in breast cancer screening.

Background: AI systems are currently being introduced in breast cancer screening with promising results, especially with respect to reduction of radiologist workload, but also increased sensitivity at constant or reduced recall rates. Similarly to other technology, solid quality assurance programs for AI systems are needed to guarantee a good and constant performance. Current mammography imaging QA workflows with technical phantoms are not adapted to test AI systems. Previously, baseline variations of the output of a cancer detection AI system with repeated imaging of an anthropomorphic breast phantom under constant imaging conditions was investigated. To establish reliable QC of AI, these baseline variations warrant a more thorough examination of AI output at repeated imaging both for digital mammography (DM) and digital breast tomosynthesis (DBT).

Methods: An anthropomorphic, physical 3D breast phantom with a simulated lesion was repeatedly imaged using automatic exposure control (AEC) with both DM and DBT. AEC imaging was done using three different mammography systems: Siemens Mammomat B.brilliant, GE Senographe Pristina and Hologic Selenia Dimensions. The difference in central tendency and variance between DM and DBT were examined using Mann-Whitney U tests and Levene's test. Furthermore, the tube voltage and tube loading of the Siemens Mammomat B.brilliant were varied, based on exposure settings determined by the AEC system. For each set of exposure parameters, repeated imaging was performed. The images were analysed using an AI-based cancer detection system. Correlations between exposure settings and the AI cancer-risk assessment were studied and results for DM and DBT were compared.

Results: The results when varying exposure settings generally point towards a significant increase of the central tendency of region scores as both tube loading and tube voltage increase, for all three systems.

Furthermore, the comparison of DM and DBT using AEC showed a significant difference in central tendency of AI risk scores (Mann-Whitney U test, $\alpha=0.05$), the significance held for all systems. The variance of the AI risk score when comparing DM and DBT using AEC was significantly different for the Siemens and Hologic systems, while no significant difference in variance was found for the risk score of the GE system (Levene's test, $\alpha=0.05$).

Conclusion: The results point towards higher repeatability of AI risk scores when using DBT imaging compared to DM. The results also point towards a higher sensitivity for DBT imaging than DM.

Abbreviations

2D – Two dimensional

3D – Three dimensional

AEC – Automatic exposure control

AGD – Average glandular dose

AI – Artificial intelligence

CC - Craniocaudal

DBT – Digital breast tomosynthesis

DM – Digital mammography

DRL – Diagnostic reference level

ECIBC – European Commission Initiative on Breast Cancer

mA - milliampere

mAs – milliampere-second

mGy - milliGray

MLO – Mediolateral oblique

SNR – Signal-to-noise ratio

Table of Contents

Populärvetenskaplig Sammanfattning 1
Abstract 2
Abbreviations 3
Introduction 5
Background 5
The effect of different imaging parameters on image quality 6
Materials and methods 8
3D breast phantom 8
Mammography equipment 9
AI system 10
Method 12
Image Acquisition 12
Data Analysis 12
Results 13
AEC Imaging Using Three Different Mammography Systems 13
Varying imaging Parameters of the Siemens Mammomat B.brilliant 16
Imaging of breast Phantom for 3 Different Manufacturers, With and Without Lesion 19
Discussion 22
AEC Imaging Using Three Different Mammography Systems 22
Varying imaging Parameters of the Siemens Mammomat B.brilliant 23
Imaging of breast Phantom for 3 Different Manufacturers, With and Without Lesion 23
Further Prospects 24
Acknowledgements 25
References 26

Introduction

AI systems are currently being introduced in breast cancer screening with promising results, especially with respect to reduction of radiologist workload, but also increased sensitivity at constant or reduced recall rates(1-3). Similarly to other technology, solid quality assurance programs for AI systems are needed to guarantee a good and constant performance. Current mammography imaging QA workflows with geometric phantoms are not adapted to test AI systems. Previously, baseline variations of the output of a cancer detection AI system with repeated imaging of an anthropomorphic breast phantom under constant imaging conditions was investigated(4). To establish reliable QC of AI, these baseline variations warrant a more thorough examination of AI output at repeated imaging both for digital mammography (DM) and digital breast tomosynthesis (DBT). It would also be valuable to expand the previous study to investigate the AI systems behaviour when analysing DM and DBT images acquired using mammography systems from multiple different manufacturers.

The purpose of this project was to examine how the degree of cancer suspicion evaluated by an AI system differs when examining repeated images of a realistic breast phantom acquired with DM and DBT. This project is a part of a larger project to develop quality assurance systems and methods for AI systems in image analysis in breast cancer screening.

Background

Breast cancer is the most common cancer diagnosis among women in Sweden. Between 2013 and 2023 around 7900 women were diagnosed with breast cancer annually, and around 1400 died due to the cancer(5). In Sweden, women aged between 40 and 74 years old are regularly called for breast cancer screening(6). The frequency of the screening opportunities falls between 18 and 24 months, the exact time between screening opportunities depending on the age of the woman.

In Sweden, a standard breast cancer screening examination consists of a craniocaudal (CC) and a mediolateral oblique (MLO) imaging of the breast using a 2D digital mammography (DM)(7). In CC imaging the breast is imaged from a top-down view while the breast is compressed from above and below. MLO images are a side-angle view, where the breast is compressed in an angled position.

An alternative to DM imaging is digital breast tomosynthesis (DBT). During DBT imaging, the x-ray tube is moved in an arc as the breast is compressed. Multiple 2D images are acquired at different angles and then reconstructed into a pseudo-3D volume.

In mammography systems with a possibility for DBT, the sweep angle is the total angle which the x-ray tube moves during the image acquisition. An increased sweep angle is associated with an improved depth resolution, but a decrease in in-plane resolution(8).

One of the main advantages of using DBT in a clinical screening setting is a significant increase in sensitivity for breast cancer detection(9). While one of the main drawbacks in implementing DBT as the primary tool for breast cancer screening is an increased image reading time for radiologists, which has shown to be up to double that of traditional DM examinations(10).

In Sweden, the diagnostic reference level (DRL) for a single-view mammography used for breast cancer screening corresponds to an average glandular dose (AGD) of 1.1 mGy(11). The DRL for a clinical tomosynthesis corresponds to an AGD of 1.9 mGy(11).

Currently, the European Commission Initiative on Breast Cancer (ECIBC) has a conditional recommendation for the implementation of DBT over DM in national screening programmes(12). Despite this conditional recommendation, the Swedish National Board of Health and Welfare (Socialstyrelsen) does not recommend this replacement. Their reasoning is that the overall benefit of DBT remains uncertain, as it is still unclear whether the method increases early breast cancer detection, while its implementation would also require substantial additional resources and costs(13).

The effect of different imaging parameters on image quality

To produce x-rays used for medical imaging, an x-ray tube is used. In its most basic shape, the x-ray tube consists of a filament and an anode, made of a material with a high atomic number, placed in a vacuum. As a current flows through the filament, a voltage is placed between it and the anode. When electrons are emitted from the filament they accelerate towards the anode. The interaction of accelerated electrons with the anode generates bremsstrahlung which is used for imaging(14).

Two of the most important parameters that can be controlled by the user of the medical imaging device is the current that flows through the filament (the tube loading) and the voltage between the filament and anode (the tube voltage). The tube loading is usually measured in mAs (milliampere-seconds) while the tube voltage is measured in kV (kilovolt).

An increase in the tube loading means that more electrons are released from the filament and by extension an increased amount of bremsstrahlung is generated. This bremsstrahlung is higher in intensity but other factors such as the energy spectrum is unchanged. This leads to an increase in the signal to noise ratio (SNR) but leaves other aspects of the image unchanged. An increase in tube voltage however accelerates the electrons that produce the bremsstrahlung more intensely, which creates bremsstrahlung with a higher energy. The higher energy of the bremsstrahlung leads to a smaller difference in attenuation between different types of tissue in the breast, leading to lower contrast in the final image.

The spectral composition of the x-ray beam used for imaging can also be influenced by the choice of anode material as well as usage of external filtering. The production of

bremsstrahlung is proportional to the atomic number of the anode material(15). The composition of the x-ray spectrum can also be modified by passing the beam through a filter material before it reaches the patient's breast. This process attenuates low-energy photons, which would otherwise be absorbed within the breast tissue and contribute to radiation dose without improving image quality. Consequently, filtration reduces the absorbed organ dose while maintaining the diagnostically useful portion of the spectrum. Similarly to the anode, the material of the filter will also change the effect of filtering, as the bremsstrahlung is more efficiently absorbed by filters with a higher atomic number(15). The optimal anode and filter materials will depend on several factors, such as detector material and breast thickness. For example, a molybdenum anode combined with a molybdenum filter give lower doses for 20mm breasts, while a tungsten anode paired with a rhodium filter gives lower doses for thicker breasts(16).

Materials and methods

3D breast phantom

The object that was imaged throughout this project was a 3D breast phantom (59-01, PhantomX GmBH, Berlin, Germany). This phantom, shown in Figure 1, consists of four slabs each with a thickness of 9mm which are held together using a magnetic mount to simulate a 36mm thick breast. The central two slabs can be replaced with an alternative pair of slabs to introduce a spiculated mass as to simulate a soft tissue lesion(17). Figure 1 shows a mammography taken where the slabs including the lesion are used as well as a mammogram where the slabs with no lesion present are used.

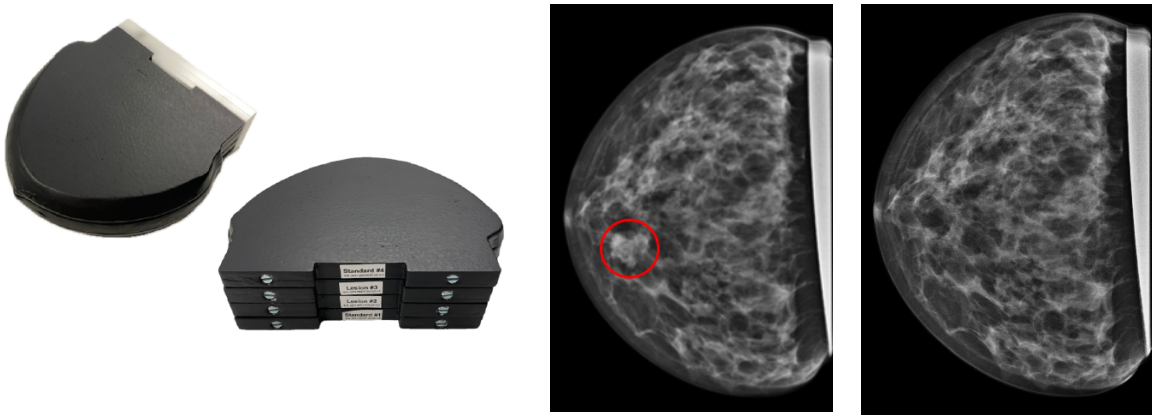


Figure 1. 3D anthropomorphic phantom (left) and mammograms of the phantom with lesion marked with a red circle (middle) and a mammogram using slabs without lesion (right)

The thickness of the breast phantom is below the mean breast thickness measured from screening patients in Malmö, which is 49.4±8.5mm. However, mammograms of the phantom have been evaluated by radiologists in Malmö, who concluded that the images generated from phantom imaging exhibit a realistic appearance. It is worth noting that the phantom is constructed to resemble a breast undergoing a CC mammography, meaning that MLO imaging is not possible.

Mammography equipment

Figure 2. The three mammography systems used in this study. The Siemens Mammomat B.brilliant (left), GE Senographe Pristina (middle) and the Hologic Selenia Dimensions (right)

In this study three different mammography systems were used: Siemens Mammomat B.brilliant (Siemens Healthineers AG, Forchheim, Germany), Ge Senographe Pristina (GE HealthCare, Chicago, Illinois, USA) and Hologic Selenia Dimensions (Hologic, Inc., Marlborough, Massachusetts, USA).

It is worth noting that all three systems perform DBT imaging with different sweep angles. The Siemens Mammomat B.brilliant provides a total sweep angle of 50°, the GE system provides a sweep angle of 25° while the Hologic system offers the smallest sweep angle, 15°.

When imaging the phantom using automatic exposure control (AEC), each system determines imaging parameters based on the attenuation of the breast, Table 1. The anode and filter materials used by each mammography machine for each imaging modality is included in Table 2.

Table 1. Parameters determined by the automatic exposure control (AEC) of the three mammography systems.

Manufacturer	Modality	Tube Voltage	Tube Loading
Siemens	DM	26kV	80mAs or 90mAs
	DBT	26kV	110mAs
GE	DM	34kV	20mAs
	DBT	34kV	26mAs
Hologic	DM	29kV	64mAs
	DBT	28kV	63mAs

Table 2. Anode and Filter material used in this study by each of the three mammography systems.

Manufacturer	Modality	Anode Material	Filter Material
Siemens	DM	Tungsten	Aluminium
	DBT	Tungsten	Aluminium
GE	DM	Rhodium	Silver
	DBT	Rhodium	Silver
Hologic	DM	Tungsten	Rhodium
	DBT	Tungsten	Aluminium

AI system

The AI system used in this project was Transpara (version 2.1.1, ScreenPoint Medical, Nijmegen, Netherlands). Transpara is an FDA cleared and CE certified AI analysis system(18, 19). As of version 1.4.0 this system is trained, validated and tested on a database of 9000 confirmed cancer cases and 180000 non-cancerous mammograms(20). It is unknown if later versions of Transpara have expanded the training dataset. However, as Transpara is a commercial AI system, many aspects of its internal operations remain undisclosed. Transpara can process mammography images. If these images belong to the same examination (such as in the case of a regular screening exam consisting of one CC and one MLO mammogram of each breast) then Transpara can examine the images together.

After an image (or series of images) has been examined by Transpara, two types of scores are assigned to the examination. These two scores are called the region score and the Transpara score (sometimes the Transpara score is also called the exam score). The region score is lesion specific and indicates the likelihood that a cancer is present at a specific location on a scale from 1 to 100. It is possible for one examination to have multiple region scores. The Transpara

score is exam-based and represent the likelihood of a cancer being present at any place in the full examination. The Transpara score is calculated using the finding with the highest region score. The Transpara score is an integer with a value of 1 to 10. However, Transpara can also output a “raw” Transpara score which is represented as a floating-point value. The regular Transpara score is simply the “raw” score rounded up.

Method

Image Acquisition

Firstly, both DM and DBT images of the phantom with the simulated lesion were collected using AEC settings (tube voltage and tube loading) was performed with the three mammography systems used in this project. For each combination of imaging modality (DM and DBT) and mammography system, 10 images were collected, except for DM images using the Siemens Mammomat B.brilliant, where fluctuations in the tube loading warranted a higher number of images so in total 17 images were collected.

Manual image acquisition was done using the Siemens Mammomat B.brilliant by varying the exposure parameters (tube loading and tube voltage), with the settings chosen by the automatic exposure control used as the basis. The exact parameters used are shown in Table 3. Ten images were acquired for each combination of exposure parameters and image modality. At 35kV only images at tube loadings of 50mAs, 80mAs, and 110mAs were used, as tube loadings of 140mAs and 200mAs resulted in overexposed images.

Table 3. Acquisition parameters used for the Siemens Mammomat B.brilliant.

Parameter	Values Used
Tube voltage (kV)	23, 26, 30, 35
Tube loading (mAs)	50, 80, 110, 140, 200

Finally, using the Siemens Mammomat B.brilliant and the GE Senographe Pristina systems, a further image series of the phantom with lesion present was collected at a tube voltage of 26kV and 5 different tube loadings, 50mAs, 80mAs, 110mAs, 140mAs and 200mAs. 10 images were collected for each tube loading. This was performed using both DM and DBT imaging. Due to restrictions of the Hologic Selenia Dimensions, imaging above 100mAs could not be performed. Instead, image series at tube loadings of 25mAs, 50mAs, 80mAs and 100mAs were collected. An image series with the same settings was collected where the lesion had been removed from the phantom. This was done using all 3 mammography systems.

Data Analysis

All DM images and DBT volumes acquired in this project were analysed using Transpara. The two scores (Transpara score and region score) were noted. Normality of the residuals of the output was not assumed and so, when selecting tests for statistical analysis, non-parametric tests were selected. Linear regression analysis was performed using the scipy package(21) (Version 1.13.1) in python. Other statistical analysis was performed using the Pingouin package(22) (Version 0.5.5) in python. For all statistical tests, a significance level of $\alpha=0.05$ was used.

For the AEC imaging, the highest region score and the Transpara score for each image were noted. The difference of central tendency between DM and DBT imaging for the one mammography system was investigated using a Mann-Whitney U-test, while the

difference in variance between the AI analysis of the two imaging modalities was investigated using Levene's test. Furthermore, the location of all detected lesions in the AEC images was visualised. This was done by creating a 2D histogram of lesion locations where the intensity of each bin is equal to the sum of regions scores at each coordinate. As the pixel width and height of the breast image matrices are in the order of 10^3 (exact value varies depending on the mammography system), a gaussian filter was used to make visualisation easier. The histogram could then be overlayed on an example image.

The results of AI analysis of imaging when varying exposure parameters using the Siemens Mammomat B.brilliant were analysed using linear regression analysis. This analysis was performed exclusively on the region score, not on the Transpara score. This was done as the point of interest was how the AI system evaluates the specific lesion, not the examination. Linear regression analysis was performed with region score as a function of the tube loading (where the tube voltage was kept constant) and as a function of tube voltage (where tube loading was kept constant).

Finally, the linear regression analysis was performed for the datasets collected when varying the tube loading for all systems, both with and without lesion present. Here, the change in Transpara score as a function of the tube loading was analysed, not the region score. The Transpara score was analysed here as half of the images were taken without a lesion present, if there is no lesion present then there is no point of comparison. This was done because the region score is intended to correspond to a lesion and therefore analysing it in images without the lesion would be inappropriate.

Results

AEC Imaging Using Three Different Mammography Systems

Figure 3 shows the raw Transpara score and the highest region score from analysing the images using Transpara. Note the difference in spread of the data when comparing the Transpara score between imaging modalities for the Siemens B.brilliant and Hologic Selenia Dimensions. Also note the substantial differences in central tendency observed when comparing both the Transpara score and the region score between DM and DBT across all systems.

Figure 4 shows the location of all lesions identified by Transpara when analysing the DM and DBT images when using the AEC settings. One should observe the large variation in lesion locations identified by the AI when analysing the DM images. Also note that, when using DBT, the AI exclusively localizes the lesion at the site of the artificial soft-tissue lesion in the phantom and does not identify any additional lesion locations.

Table 4, Table 5 and Table 6 show the results of the statistical tests between the DM and DBT results for the Siemens, GE and Hologic systems respectively.

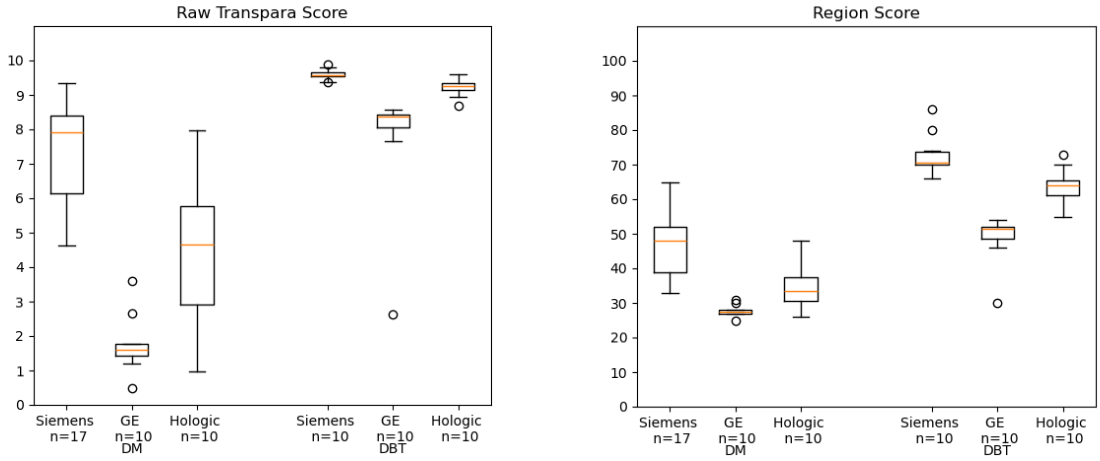


Figure 3. Raw Transpara score (left) and region score (right) of the images of the 3D breast phantom using the Siemens Mammomat B.brilliant, GE Senographe Pristina and the Hologic Selenia Dimensions. Tube loading and tube voltage were determined by the AEC. The boxplots are grouped by imaging modality, DM on the left and DBT on the right.

For the Siemens Mammomat B.brilliant, the mean value of the raw Transpara score was 9.59±0.16 for DBT and 7.40±1.31 for DM. The mean value of the region scores was 72.6±5.89 for DBT and 46.6±8.28 for DM.

For the GE system, the mean value of the raw Transpara score was 7.71±1.71 for DBT and 1.75±0.80 for DM. The mean value of the highest region scores was 48.7±6.60 for DBT and 27.8±1.60 for DM.

For the Hologic system, the mean value of the raw Transpara score was 9.22±0.26 for DBT and 4.42±2.04 for DM. The mean value of the highest region scores was 63.7±5.00 for DBT and 34.5±6.09 for DM.

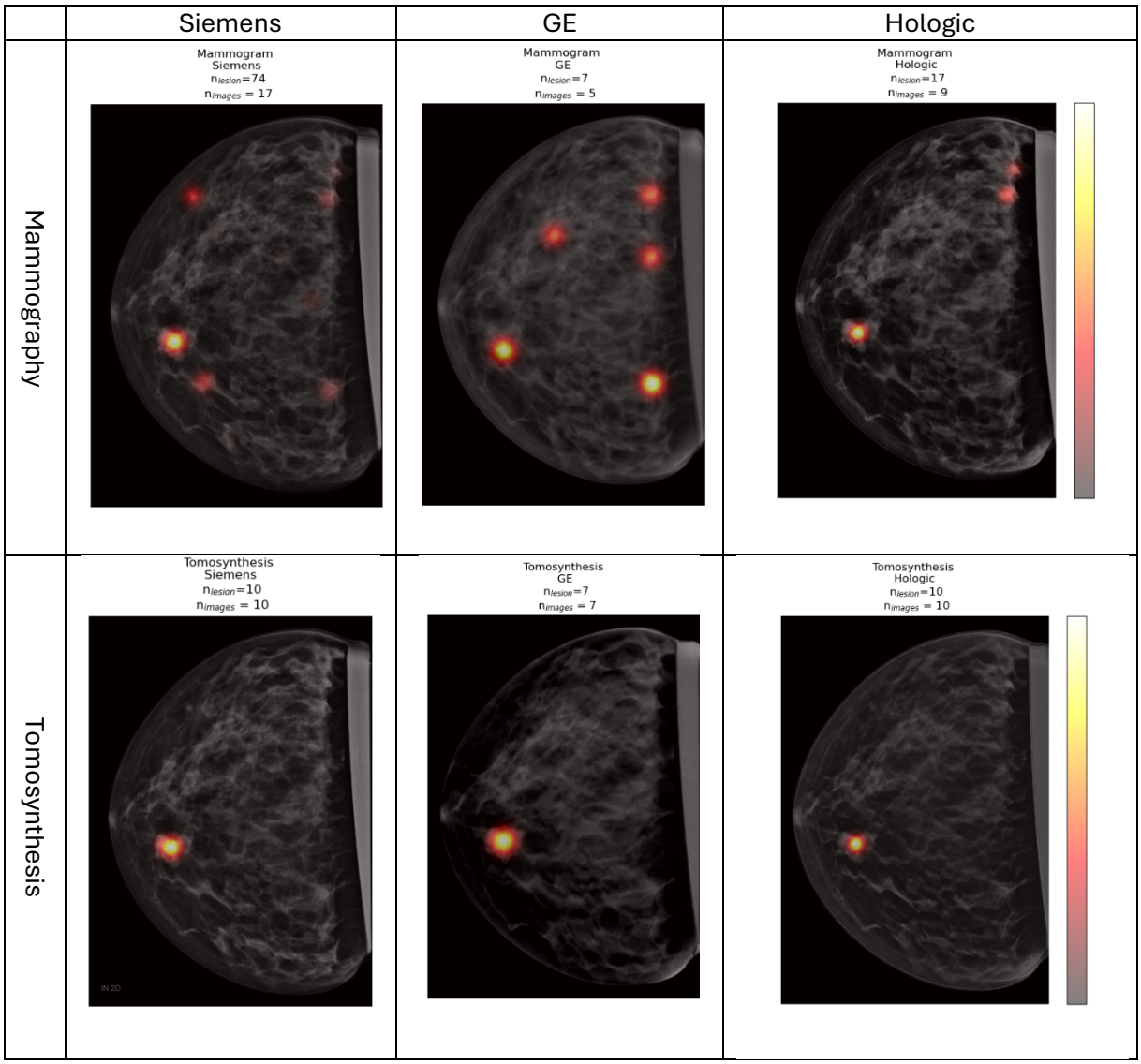


Figure 4. Lesion locations as identified by Transpara when imaging using AEC. n_{lesion} is the number of lesions in total that were identified across the number of images, n_{images}. Note that in all cases (except for DM images in the Siemens Mammomat B.brilliant), 10 images in total were analysed. This means that in cases where n_{images}<10, there are images where no lesion was identified.

Table 4.Results of statistical tests comparing the distribution of Transpara score and Region score between DM and DBT images for the Siemens Mammomat B.brilliant.

Manufacturer	Compared Score	Test	P-value
Siemens	Raw Transpara score	Mann-Whitney U	<0.001
	Raw Transpara score	Levene	0.001
	Region Score	Mann-Whitney U	<0.001
	Region Score	Levene	0.156

Table 5.Results of statistical tests comparing the distribution of Transpara score and Region score between DM and DBT images for the GE Senographe Pristina.

Manufacturer	Compared Score	Test	P-value
GE	Raw Transpara score	Mann-Whitney U	<0.001
	Raw Transpara score	Levene	0.75
	Region Score	Mann-Whitney U	<0.001
	Region Score	Levene	0.25

Table 6.Results of statistical tests comparing the distribution of Transpara score and Region score between DM and DBT images for the Hologic Selenia Dimensions.

Manufacturer	Compared Score	Test	P-value
Hologic	Raw Transpara score	Mann-Whitney U	<0.001
	Raw Transpara score	Levene	0.002
	Region Score	Mann-Whitney U	<0.001
	Region Score	Levene	0.57

Varying imaging Parameters of the Siemens Mammomat B.brilliant

Figure 5 shows the AI analysis of DM images using four different tube voltages, varied between the mammography system's maximum and minimum settings, and five different tube loadings. These are plotted over the average organ dose. In the results of the Transpara score determined by AI analysis of DM imaging, one sees a steep rise as organ dose increases, which eventually plateaus as the Transpara score nears 10. Figure 6 shows the results of the AI analysis of the corresponding DBTs.

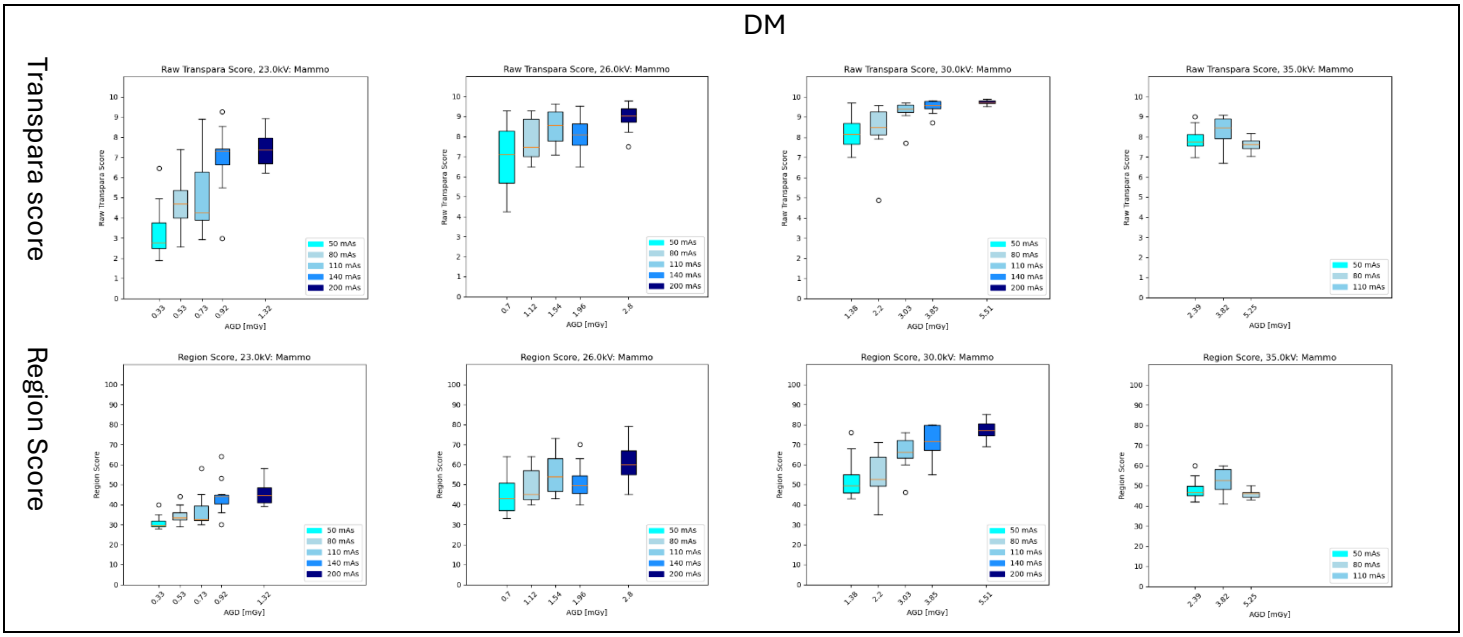


Figure 5. Raw Transpara score (top) and Region score (bottom) of DM images of the 3D breast phantom using varying imaging conditions for the Siemens Mammomat B.brilliant.

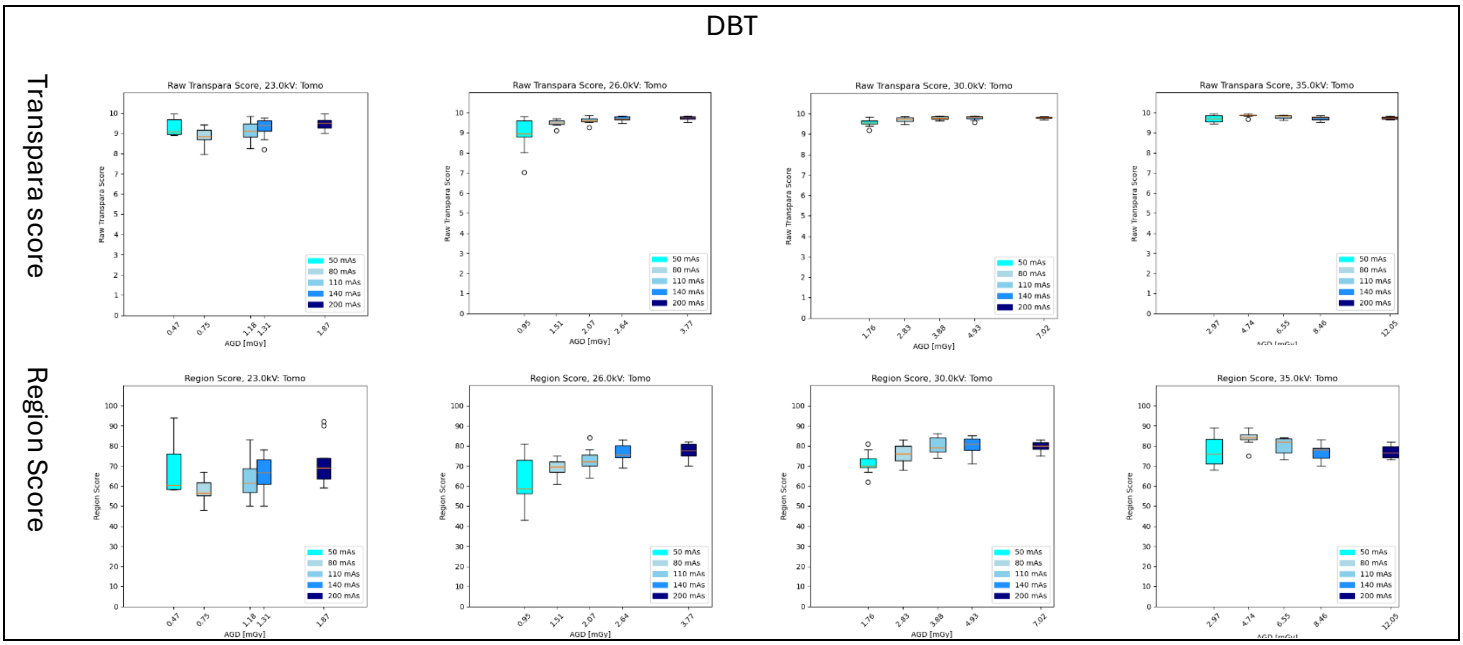


Figure 6.Raw Transpara score (top) and Region score (bottom) of DBT images of the 3D breast phantom using varying imaging conditions for the Siemens Mammomat B.brilliant.

Table 7 shows the results of the linear regression analysis of the region score of the DM and DBT images as the tube loading is varied while tube voltage is kept constant for the Siemens Mammomat B.brilliant. Table 8 shows the results of linear regression analysis of region scores when the tube loading is kept constant and tube loading is varied.

*Table 7.Results of linear regression analysis comparing the distribution of region score between DM and DBT images for the Siemens Mammomat B.brilliant when tube voltage is kept constant. (*Significant at α=0.05)*

Modality	Imaging Conditions	Linear Regression Coefficient	P-value	R^2
DM	kV = 23 mAs Varied	3.78*	<0.001	0.381
	kV = 26 mAs Varied	2.21*	<0.001	0.207
	kV = 30 mAs Varied	2.93*	<0.001	0.508
	kV = 35 mAs Varied	-0.77	0.351	0.031
DBT	kV = 23 mAs Varied	1.01	0.115	0.048
	kV = 26 mAs Varied	3.35*	<0.001	0.318
	kV = 30 mAs Varied	4.74*	<0.001	0.25
	kV = 35 mAs Varied	-1.84	0.177	0.037

*Table 8.Results of linear regression analysis comparing the distribution of region score between DM and DBT images for the Siemens Mammomat B.brilliant when tube loading is kept constant. (*Significant at α=0.05)*

Modality	Imaging Conditions	Linear Regression Coefficient	P-value	R^2
DM	kV Varied mAs = 50	0.22*	<0.001	0.302
	kV Varied mAs = 80	0.22*	<0.001	0.291
	kV Varied mAs =110	0.07	0.153	0.053
	kV Varied mAs = 140	0.16*	<0.001	0.625
	kV Varied mAs = 200	0.17*	<0.001	0.772
DBT	kV Varied mAs = 50	0.15*	0.019	0.137
	kV Varied mAs = 80	0.37*	<0.001	0.773
	kV Varied mAs =110	0.33*	<0.001	0.452
	kV Varied mAs = 140	0.27*	0.002	0.218
	kV Varied mAs = 200	0.18	0.090	0.074

Imaging of breast Phantom for 3 Different Manufacturers, With and Without Lesion

Figure 7 shows the AI analysis of DM images and the three different mammography systems. Tube voltage is kept constant at 26kV. The results are plotted over the average AGD as calculated by the system. Figure 8 shows the results of the AI analysis using the same exposure settings and the two central slabs replaced with the two corresponding slabs with no lesion.

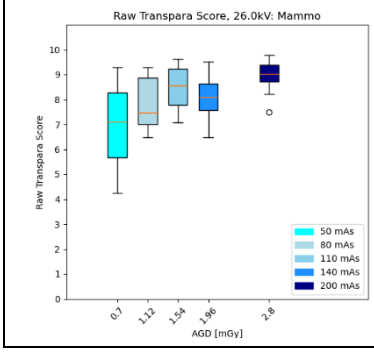
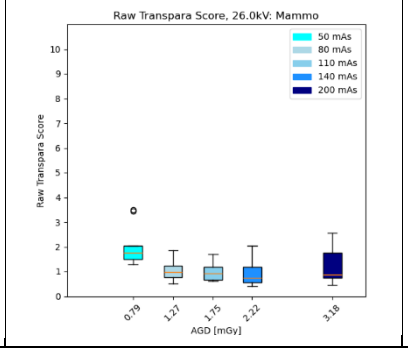
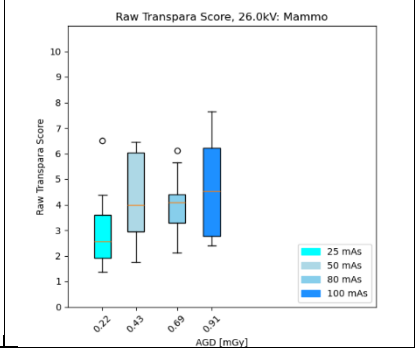
DM – Lesion Present		
Siemens	GE	Hologic
		

Figure 7. Raw Transpara score as a function of dose for DM images using 3 different mammography systems.

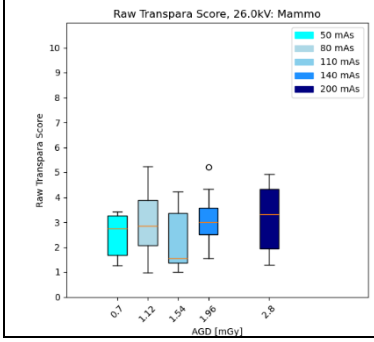
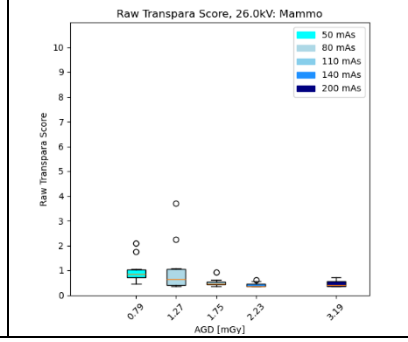
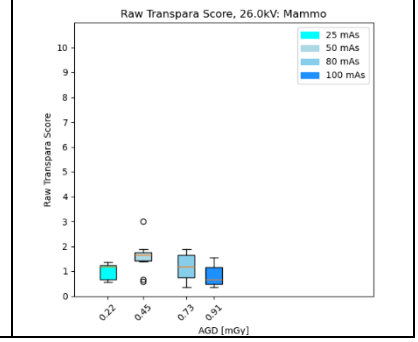
DM – Lesion Not Present		
Siemens	GE	Hologic
		

Figure 8. Raw Transpara score as a function of dose for DM images for the 3 mammography systems when lesion has been removed from the phantom.

Figure 9 shows the result of AI analysis when using DBT imaging for three mammography systems. Tube voltage is kept constant at 26kV. Figure 10shows the result of AI analysis when the lesion is removed but all other exposure settings are kept the same.

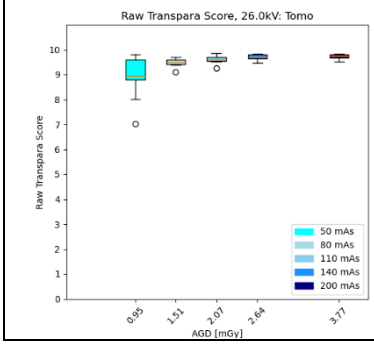
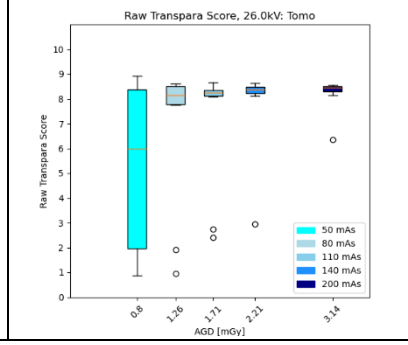
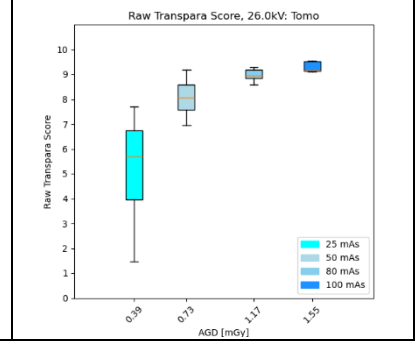
DBT – Lesion Present		
Siemens	GE	Hologic
		

Figure 9. Raw Transpara score as a function of AGD for DBT volumes when for the 3 mammography systems.

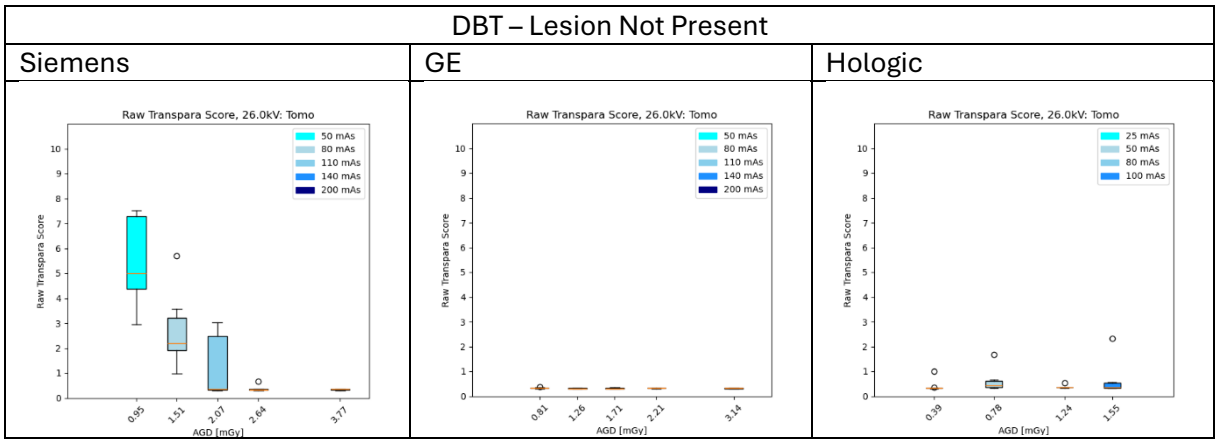


Figure 10. Raw Transpara score when analysing DBT volumes as a function of the AGD for the 3 mammography systems. Lesion has been removed from the phantom.

Table 9 shows the results of the linear regression analysis of the region score of the DM and DBT AI image analysis shown in Figure 7 and Figure 9 as the tube loading is varied and a lesion is present in the phantom. Table 10 shows the results of linear regression analysis when the lesion is removed.

*Table 9.Results of linear regression analysis of the Transpara score when varying tube loading for the three different mammography systems with the lesion present in the phantom. (*Significant at α=0.05)*

Modality	Manufacturer	Imaging Conditions	Linear Regression Coefficient	P-value	R^2
DM	Siemens	kV = 26 mAs Varied Lesion Present	20.04*	<0.001	0.224
	GE	kV = 26 mAs Varied Lesion Present	-20.20	0.053	0.076
	Hologic	kV = 26 mAs Varied Lesion Present	5.21	0.052	0.096
DBT	Siemens	kV = 26 mAs Varied Lesion Present	4.32	0.473	0.011
	GE	kV = 26 mAs Varied Lesion Present	12.30	0.245	0.028
	Hologic	kV = 26 mAs Varied Lesion Present	0.77	0.78	0.002

*Table 10. Results of linear regression analysis of the Transpara score when varying tube loading for the three different mammography systems and no lesion present in the phantom. (*Significant at α=0.05)*

Modality	Manufacturer	Imaging Conditions	Linear Regression Coefficient	P-value	R^2
DM	Siemens	kV = 26 mAs Varied No Lesion Present	7.66	0.203	0.034
	GE	kV = 26 mAs Varied No Lesion Present	-33.86*	0.005	0.151
	Hologic	kV = 26 mAs Varied No Lesion Present	-8.57	0.291	0.029
DBT	Siemens	kV = 26 mAs Varied No Lesion Present	6.84	0.257	0.027
	GE	kV = 26 mAs Varied No Lesion Present	-14.81	0.238	0.029
	Hologic	kV = 26 mAs Varied No Lesion Present	9.47	0.242	0.036

Discussion

A factor that permeates all results in this study is that the results of AI analysis under identical imaging conditions still vary. One would expect that when imaging conditions are kept constant, the output would also remain constant. The reason for this is unknown, as we do not have complete insight into the structure of the AI. However, it is likely that this is a consequence of variation in the noise of the images.

AEC Imaging Using Three Different Mammography Systems

The results in Figure 3 and Table 4 -Table 5 show that when using the Siemens and the Hologic systems, the repeatability of the Transpara analysis is higher for DBT images compared to DM. Furthermore, the scores for DBT images are significantly higher than for DM images regardless of the manufacturer. As the breast phantom has a simulated lesion present, a high Transpara score is expected. Figure 4 further strengthens the reliability of DBT as an imaging modality when using AI analysis. These results indicate that lesion localization is substantially more robust in DBT than in DM imaging. For DBT, the AI system consistently identifies only the intended soft-tissue lesion as suspicious, whereas in DM it frequently highlights additional non-malignant tissue regions as suspicious findings. In the DBT cases where the intended lesion is not identified as suspicious, no other tissue is suspected either.

However, a recurring problem in this project has been the lack of a clear ground truth. Transpara is trained using real patient cases meaning that imaging of a breast phantom may not replicate the results from analysing DM or DBT images of real breasts. Even though radiologists in Malmö have evaluated the phantom as realistic, their assessment may differ from the AI. The AI may correctly determine that the soft-tissue lesion within the phantom is not malignant, as the phantom lesion is not true cancerous tissue. It should be noted that the variance of the output is expected to remain unaffected regardless of whether the AI system identifies a lesion or not.

Varying imaging Parameters of the Siemens Mammomat B.brilliant

Table 7 and Table 8 show that region score often has a significant increase with an increase in tube loading och tube voltage. However, the variance in the data is only limitedly explained by the linear regression, as can be seen by the R^2-values for the significant results being below 0.50 for all but 4 regressions. This is likely a consequence of the relationship not actually being linear. Figure 5 and Figure 6 seem to trend towards a plateau as the AGD increases. If one wishes to completely describe the behaviour of the AI system, then one would have to assume a non-linear relationship, as the region score cannot go above 100 or below 1 which any linear regression with a linear regression coefficient not equal to 0 would eventually cross.

Imaging of breast Phantom for 3 Different Manufacturers, With and Without Lesion

Table 9 show the results of linear regression analysis of the Transpara score when varying tube loading for the three different mammography systems while keeping the tube voltage constant. While only the linear regression of the AI analysis of the DM images from the Siemens Mammomat B.brilliant shows significance, the analysis of the results from the GE Senographe Pristina and the Hologic Selenia Dimensions seem to be approaching significance, with p-values of 0.053 and 0.052 respectively. While not significant, the results when using the Hologic Selenia Dimensions seems to point towards an increase in Transpara score as the organ dose increases. However, these observations suggest that a linear regression model may not be sufficient to characterize the behaviour of the AI. Furthermore, a larger sampling of tube loadings and larger sample sizes for each data point would provide a more representative characterisation of the AI behaviour. Until such characterisation has been performed, it is difficult to propose alternative analytical approaches. This is due to the limitations arising from the AI system's behaviour being unknown.

Table 10 shows that when the slabs of the phantom where no lesion is present are used, no significance is observed, except for the AI analysis of the DM images obtained with the GE Senographe Pristina. When observing Figure 8 and Figure 10, this seems to come because of the Transpara score being low at low tube loadings. If one assumes the phantom to be an accurate representation of a human breast, then this would point towards the variance of the AI output not introducing false positives.

Care should be taken when extrapolating the results of the AI analyses of the DM and DBT imaging from the GE Senographe Pristina in Table 9 and Table 10, as these were acquired using a tube loading of 26kV, well outside the AEC setting of 34kV shown in Table 1.

Further Prospects

This project has partially shown how the Transpara system behaves as imaging conditions vary. However, this study is not fully comprehensive. It would be interesting to collect full imaging series, like those described in Table 3 for all mammography systems.

The scope of the project could be further broadened by including a larger number of mammography systems. This could include not only systems from additional manufacturers, but also multiple models of mammography systems from manufacturers that are represented in the present study.

The number of AI systems could also be expanded to investigate if the behaviours described in this project are applicable to AIs in general or if it is exclusively the behaviour of Transpara.

Finally, it would be interesting to see if the behaviours described in this study are consistent when AI analysis applied to repeated imaging of patient data. If this could be shown, then quality control using anatomical phantoms would be solidified as a representative method when studying the behaviour of AI systems.

Acknowledgements

Firstly, I would like to thank my supervisors **Anders Tingberg** and **Magnus Dustler** for their support and constant feedback during the project. Without both their expertise and their personal engagement would not have been possible.

Furthermore, the project would not be doable without **Lina Alström, Viktor Dahlbom** and **Anna Bjerkén**. Their knowledge of various subjects, such as the usage of the GE Senographe Pristina and how the AI analysis program is run, were essential to the completion of the project.

I would also like to thank **Unilabs SUS Malmö** for allowing me to use their mammography systems for data collection during this project. In a similar vein I would also like to thank two medical physicists at Karolinska Universitetssjukhuset, **Jörgen Scherp Nilsson & Markus Hulthén** for allowing me to come visit the hospital and broaden the scope of the project.

Another group I would like to thank all the students who were completing their thesis (Bachelors and Masters) for helping me keep a positive attitude during the completion of the project. I would like to thank **Emma Kristoffersson** and **Matilda Bilén** whose intense interest for the winter Olympics and puzzles kept spirits high.

I also want to extend my appreciation to every member of the **LUCI** Group their openness and the enthusiasm for my project was something that allowed me to increase the quality of my work and to gain a greater understanding of the context in which my project exists.

Finally, I would like to thank my friends and family for the kindness and support, which has been with me even before I began studying for my Master of Science.

References

1. Dahlblom V, Andersson I, Lang K, Tingberg A, Zackrisson S, Dustler M. Artificial Intelligence Detection of Missed Cancers at Digital Mammography That Were Detected at Digital Breast Tomosynthesis. Radiol Artif Intell. 2021;3(6):e200299.
2. Dembrower K, Crippa A, Colon E, Eklund M, Strand F, ScreenTrust CADTC. Artificial intelligence for breast cancer detection in screening mammography in Sweden: a prospective, population-based, paired-reader, non-inferiority study. Lancet Digit Health. 2023;5(10):e703–e11.
3. Hernstrom V, Josefsson V, Sartor H, Schmidt D, Larsson AM, Hofvind S, et al. Screening performance and characteristics of breast cancer detected in the Mammography Screening with Artificial Intelligence trial (MASAI): a randomised, controlled, parallel-group, non-inferiority, single-blinded, screening accuracy study. Lancet Digit Health. 2025;7(3):e175–e83.
4. Alström L, Bjerkén A, Dahlblom V, Timberg P, Dustler M, Tingberg A. Assessing the robustness of AI lesion risk scores at different exposure settings using an anthropomorphic breast phantom. Radiation Protection Dosimetry. In Press.
5. Socialstyrelsen. Statistik om bröstcancer. In: Socialstyrelsen, editor. 2023.
6. Socialstyrelsen. Bröstcancer – screening med mammografi 2023 [Available from: https://www.socialstyrelsen.se/kunskapsstod-och-regler/regler-och-riktlinjer/nationella-screeningprogram/slutliga-rekommendationer/brostcancer/.
7. Vårdguiden. Regelbunden mammografi för att upptäcka bröstcancer 2025 [2026-05-29]. Available from: https://www.1177.se/undersokning-behandling/undersokningar-och-provtagning/upptack-och-forebygg-cancer/regelbunden-mammografi-for-att-upptacka-brostcancer.
8. Lai YC, Ray KM, Mainprize JG, Kelil T, Joe BN. Digital Breast Tomosynthesis: Technique and Common Artifacts. J Breast Imaging. 2020;2(6):615–28.
9. Zackrisson S, Lang K, Rosso A, Johnson K, Dustler M, Fornvik D, et al. One-view breast tomosynthesis versus two-view mammography in the Malmo Breast Tomosynthesis Screening Trial (MBTST): a prospective, population-based, diagnostic accuracy study. Lancet Oncol. 2018;19(11):1493–503.
10. G JWP, Darker I, J JJ, Satchithananda K, Sharma N, Valencia A, et al. How long does it take to read a mammogram? Investigating the reading time of digital breast tomosynthesis and digital mammography. Eur J Radiol. 2024;177:111535.
11. Strålsäkerhetsmyndigheten. SSMFS 2018:5 Strålsäkerhetsmyndighetens föreskrifter och allmänna råd om medicinska exponeringar (konsoliderad version). Stockholm: Strålsäkerhetsmyndigheten; 2018.
12. (ECIBC) ECIoBC. European breast cancer guidelines [Available from: https://cancer-screening-and-care.jrc.ec.europa.eu/en/ecibc/european-breast-cancer-guidelines?topic=62&usertype=60#guideline.
13. Socialstyrelsen. Screening för bröstcancer – Socialstyrelsens rekommendation – Slutversion 2023. Stockholm: Socialstyrelsen; 2023. Report No.: 2023-5-8564.
14. Dance DR, Christofides S, Maidment ADA, McLean ID, Ng K-H, International Atomic Energy A. Diagnostic radiology physics : a handbook for teachers and students. 1st ed. Vienna: International Atomic Energy Agency; 2014.
15. Podgoršak EB, SpringerLink. Radiation Physics for Medical Physicists. 3rd 2016. ed. Cham: Springer International Publishing : Imprint: Springer; 2016.
16. Dance DR, Thilander AK, Sandborg M, Skinner CL, Castellano IA, Carlsson GA. Influence of anode/filter material and tube potential on contrast, signal-to-noise ratio and average absorbed dose in mammography: a Monte Carlo study. Br J Radiol. 2000;73(874):1056–67.
17. PhantomX. [2026-05-27]. Available from: https://phantomx.de/.

18. Administration USFaD. 510(k) Premarket Notification K241831 2024 [updated 2026–06–08. Available from: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K241831.
19. Medical S. Transpara Regulatory Clearances 2025 [Available from: https://screenpoint-medical.com/insights/transpara-regulatory-clearances.
20. 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-J Natl Cancer I. 2019;111(9):916–22.
21. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020;17(3):261–72.
22. Vallat R. Pingouin: statistics in Python. J Open Source Softw. 2018;3(31):1026.