



LUND
UNIVERSITY

Master of Science Thesis
VT2019

Artificial intelligence together with mechanical imaging in mammography

Anna Bejnö

Supervisors

Magnus Dustler and Anders Tingberg

This work has been conducted at
Medical Radiation Physics, Skane University Hospital, Malmö

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

Populärvetenskaplig sammanfattning

Bröstcancer är den vanligaste cancerdiagnosen för kvinnor och årligen drabbas över 9000 i Sverige. För att hitta cancer i ett tidigt stadium genomförs regelbunden screening (massundersökning) med mammografi (bröströntgen) för kvinnor i åldern 40–74. Artificiell intelligens (AI) och mekanisk avbildning, Mechanical Imaging (MI), har använts i separata studier inom mammografi för att underlätta granskningen av mammografibilderna. Teknikerna har individuellt visat stor potential, men användningen av de två teknikerna tillsammans har aldrig utvärderats.

AI-programmet som använts i den här studien är ett djupinlärningsprogram som hittar fynd utifrån misstänksamma mönster i mammografibilderna. Programmet poängsätter också fynden, vilka indikerar en nivå av misstänksamhet för bröstcancer. Mechanical Imaging är en mätning av trycket vid bröstets yta med hjälp av mekaniska sensorer då det komprimeras i samband med bildtagningen. Mätningen ger information om den underliggande vävnadens mekaniska strukturer och om det finns något styvt område i bröstet. Elakartade tumörer har påvisats vara styvare än godartade förändringar och tekniken skulle därför kunna användas som ett komplement vid screeningtillfället.

I arbetet har mammografibilder med tillhörande tryckmätningar från 46 kvinnor använts för att analysera AI-fynd tillsammans med tryckbilderna. Av dessa kvinnor hade 12 cancer, 15 hade benigna (godartade) tumörer och 19 hade diagnostiserats som friska. AI-programmet analyserar mammografibilder närmast som en radiolog. Det ger ett värde för hela mammografiundersökningen på en skala på 1–10 hur pass sannolikt det är att kvinnan har cancer, där 1 motsvarar en låg sannolikhet och 10 motsvarar hög sannolikhet för att kvinnan har cancer.

Resultatet av studien visar att det finns tendenser till att ett ökat tryck över ett AI-fynd skiljer sig mellan sjuka och friska, och att metoderna tillsammans kan hitta oberoende markörer som båda sammanfaller i bröstcancer. Användbarheten av MI och AI tillsammans har därför stor potential att kunna minska andelen återkallade kvinnor, om metoden skulle införas vid screeningtillfället. Det här är första utvärderingen av MI och AI tillsammans, och vidare studier krävs för att utvärdera metodernas fulla potential.

Abstract

Artificial intelligence (AI) and mechanical imaging (MI) have been used in separate studies in breast imaging. They have individually shown great possibilities within the field of mammography, but the use of the two techniques together have never been evaluated. The artificial intelligence application used in this work was Transpara, a deep learning convolutional neural network. It distinguishes patterns in the mammographic images and provides scores of individual findings and the whole mammographic examination, which indicates a level of suspicion for breast cancer. Mechanical imaging is a surface stress measurement, that provides information of the mechanical structure of the underlying tissue. Since malignant tumours often express a higher relative pressure compared to the surrounding tissue in the breast, mechanical imaging is comparable with palpation but could provide even more information of the mechanical structures.

The purpose of this work was to study if the combination of the two methods could be used to directly detect breast cancer. Screening images of 118 women were analysed in Transpara, and the pressure distribution measurement of the same women was obtained from a previous study on MI. For 46 cases, there was compression pressure present over the AI-findings, and these were chosen to be included in the analysis. Locations of findings with the highest level of suspicion and the corresponding locations in the pressure measurement were used to calculate the mean relative pressure over a finding. The cases were divided into three groups by diagnosis; biopsy-proven cancer, biopsy-proven benign and non-biopsied, very likely benign. The increased pressure was then compared among these three groups and the two groups of cancer and healthy, to evaluate if the increased pressure over Transpara scores of women diagnosed with cancer was different from those diagnosed as healthy. The correlation between increased pressure and Transpara score was evaluated for each group, to evaluate if the two methods found the same indications for breast cancer.

The results of this study indicated that there probably are differences in increased pressure between cases with breast cancer and healthy, but it remains to further evaluated for a larger material. A significant and relatively strong correlation between the relative pressure increase over an AI-finding and the Transpara scores was established in the group with cancer, but the other groups showed no correlation.

This study indicates that MI combined with AI can potentially be used to improve the performance of mammography screening. It suggests that AI and MI find independent markers that coincide in breast cancer. Therefore, the two methods have the potential of lowering the recall rate in mammography, but this needs to be further evaluated.

Keywords: mammography, mechanical imaging, deep learning, computer-aided detection

Abbreviations

2D - Two-dimensional

3D - Three-dimensional

AI - Artificial Intelligence

AUC - Area under the ROC Curve

CAD - Computer-assisted Detection

CC - Craniocaudal

CNN - Convolutional Neural Network

DBT - Digital Breast Tomosynthesis

DM - Digital Mammography

MI - Mechanical Imaging

MLO - Mediolateral Oblique

MLP - Multi-layer Perceptron

ROC - Receiver Operating Curve

Contents

1	Introduction.....	1
1.1	Background	1
1.2	Research question and aims	2
2	Theory	3
2.1	Mammography	3
2.2	Artificial intelligence and Deep learning	4
2.3	Transpara.....	5
2.4	Mechanical imaging.....	7
3	Materials and methods	10
3.1	Data collection	10
3.2	Suspicious lesion analysis	11
3.3	Mechanical imaging analysis	12
3.4	Statistical analysis.....	14
4	Results	16
5	Discussion	23
5.1	Future prospects.....	25
6	Conclusions.....	26
7	Acknowledgements	27
8	References.....	28

1 Introduction

1.1 Background

In Sweden alone, over 9600 women are diagnosed with breast cancer every year. This corresponds to about 15 % of all cancer diagnoses in Sweden each year and is the most common cancer diagnosed for women [1]. To address this, a high emphasis has over the years been placed on detecting breast cancer at an early stage to improve treatment outcomes. Under the responsibility of the Swedish National Board of Health and Welfare, all women aged 40-74 are offered to take part in a voluntary screening every 18-24 months [2]. The gold standard method used for screening is digital mammography, DM. A more advanced derivative of DM was developed during the late 90's, digital breast tomosynthesis, DBT [3, 4]. DBT uses the same basic technique as DM, but the x-ray tube moves around the breast in an arc (typically 15-50°). This allows the method to acquire several 2D-projections that can be used to reconstruct 3D-volumes of the breast, instead of a single 2D-projection image. This technique makes it possible for the radiologist to study individual slices of the breast and compensate for the effect of overlapping breast tissue [3].

The evaluation of the mammograms is performed by radiologists, that look for abnormalities and calcifications. Since the screening is offered every 18-24 months, the radiologists can compare the mammograms over time. With DBT the reading time increased 1-3 times compared to the reading time of DM examinations [5, 6]. Increased reading time is mainly caused by the increased number of images for the radiologists to review using DBT instead of the traditional DM [6]. To assist the radiologists, various Computer-Aided Detection, CAD, systems have been developed and evaluated. CAD systems have, however, been questionable in terms of assisting radiologists [7]. A large study with a reading of about 500,000 mammograms showed that CAD does not improve the diagnostic accuracy [8]. Ever since artificial intelligence, AI, a new version of CAD trained by deep learning algorithms got its breakthrough around 2010, researchers have been trying to implement it in medical image analysis [9]. AI systems have shown great possibilities in screening mammography for aiding radiologists. One of the AI systems that has shown promising results is Transpara, a system developed by ScreenPoint Medical, Nijmegen, the Netherlands [10]. In a study regarding

detection performance, Receiver Operating Characteristics, ROC, was used. The Area Under the ROC Curve, AUC, was in this study slightly higher for the AI system compared to that of radiologists in a non-screening material [11]. Transpara is a system that searches through the mammographic images for specific patterns, calcifications and other features that indicate cancerous tissue, and marks these abnormalities with a risk score from 0-100 (15-95) [10].

Another new technique, Mechanical Imaging, MI, has recently been developed and used in studies in breast imaging [12-14]. It is a unique technique for breasts that currently is a research field at the Department of Radiation Physics, Skåne University Hospital, Malmö. Pressure sensors measure the mechanical stress on the breast surface during compression in connection to the image acquisition. The technique could be useful since a pressure change relative to background pressure indicates that there may be a tumour present or at least some change in mechanical structure. Results from a study have shown that MI used in screening could lower the recall and the biopsy rates [13]. One of the major problems in breast cancer screening today is that most suspicious findings marked by radiologists and AI are false positives. The challenging problem, to distinguish healthy women from the ones with disease, is as important as finding those women who have the disease. The fact that MI could be used as a complement to find tumours and soft-tissue lesions is very useful and could together with AI be a powerful tool for lowering the recall rate and thus avoid worrying women unnecessarily.

1.2 Research question and aims

This study sets out to investigate ‘To what extent can MI combined with AI be used to improve the performance of mammography screening?’.

The main aims of this thesis study are to

- i. Relate masses indicated by the AI application to pressure changes relative to background pressure.
- ii. Determine what role the relation plays in clinical mammography.

2 Theory

2.1 Mammography

Digital mammography (DM) is the current standard in screening for breast cancer in Sweden [15]. In screening, two views are typically used: craniocaudal, CC, and mediolateral oblique, MLO, visualized in Figure 1. The CC view is used to better visualize tumours and abnormalities that are located in the central part of the breast whereas the MLO view includes more tissue in the upper juxtathoracic portion of the breast [16]. The CC view is usually less affected by overlapping tissue than the MLO view. This makes it rational to include both views in a DM examination.

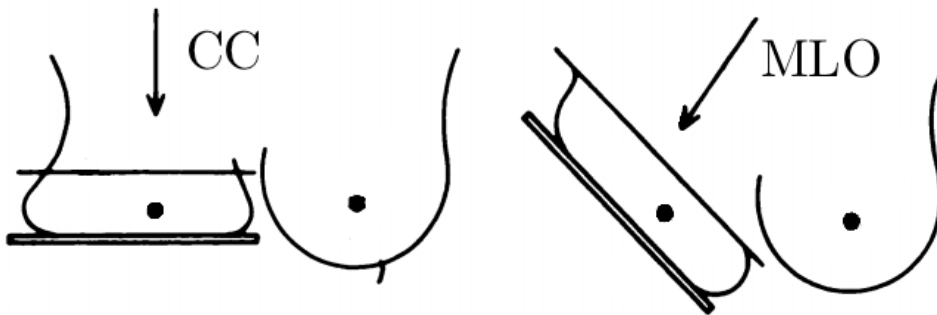


Figure 1: Mammographic views, CC and MLO, used in screening [16].

As a complement to DM, there are various other techniques used in the screening chain. This includes digital breast tomosynthesis (DBT), breast CT (BCT), ultrasound and breast magnetic resonance imaging (MRI) [17]. MRI is used for younger women with dense breasts. Additional imaging with DM in the latero-medial (LM or ML) view and an ultrasound is usually performed when a woman is recalled from the initial DM screening [17]. In some cases, a DBT scan is performed to better visualize the breast volume and to reduce the impact of overlapping tissue.

The compression of the breast during mammography and breast tomosynthesis has three major benefits; reduction of absorbed dose, improved image quality and the fixation of the breast to counteract motion blur [12, 18]. Breast compression also has negative implications, and the most severe is pain. Pain is one of the main reasons why women do not participate in screening [19]. However, over recent years the technique has developed, although a study from 2012 indicates

that the pressure distribution in mammography is not always optimal [12]. Two types of compression paddles (rigid and flexible) were compared in a study, which suggested that the flexible compression paddle is superior as the experienced pain was the same for both paddles [20]. The flexible compression paddle could then be used either with the same level of discomfort, but with improved compression or with the same level of compression but with a probable reduction of discomfort [20].

Breast tissue consists mainly of two different components; adipose and fibroglandular tissue [20]. Adipose tissue consists mainly of fat and appears dark in the mammogram while fibroglandular tissue consists of various networks of glandular and fibrous tissue which appear light in the mammogram. Women with so-called 'dense breasts' have breasts with a high proportion of fibroglandular tissue [12, 15].

The World Health Organisation, WHO, has defined a histological classification of breast tumours which on a basic level can be divided into two groups; noninvasive and invasive carcinoma [21]. Noninvasive carcinoma has two main subgroups; ductal carcinoma in situ (DCIS) and lobular carcinoma in situ (LCIS). DCIS originate from the mammary ducts and LCIS involves the intralobular ducts. The most common cases of invasive carcinoma are invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC) [21].

2.2 Artificial intelligence and Deep learning

In medical image analysis, convolutional neural networks, CNNs, are the most popular architectures for an artificial intelligence system [22]. CNN is a network that contains layers that transform the input data, e.g. images, with convolutional filters to output data, e.g. abnormalities present/absent. It is created such that patterns in images can be identified which makes CNNs very useful in image analysis. Compared to other networks such as a multi-layer perceptron, MLP, CNNs contains hidden layers called convolutional layers. The convolutional layers are layers that identify patterns by using several K kernels $\mathcal{W} = W_1, W_2, \dots, W_K$ and biases $\mathcal{B} = b_1, b_2, \dots, b_K$ that each specializes in one type of pattern, e.g. a specific shape or a calcification. Patterns are identified as the kernels are convolving over the input data, and the output from one hidden layer is then

generating a new feature map X_k^l which is subjected to an element-wise non-linear transform $t(\cdot)$ [22, 23]. The process is repeated for every hidden layer l before reaching the final output node:

$$X_k^l = t(W_k^{l-1} \cdot X^{l-1} + b_k^{l-1})$$

When a CNN is working with multiple layers, it is called a deep learning system. Another difference to MLP is that CNNs often have pooling layers which are used for dimensionality reduction. That means that pooling layers reduce the number of parameters. At the end of the network, a final layer is usually added, called a fully-connected layer, which does not use shared weights as a regular neural network layer or a MLP [22]. A basic 1D CNN is illustrated in Figure 2.

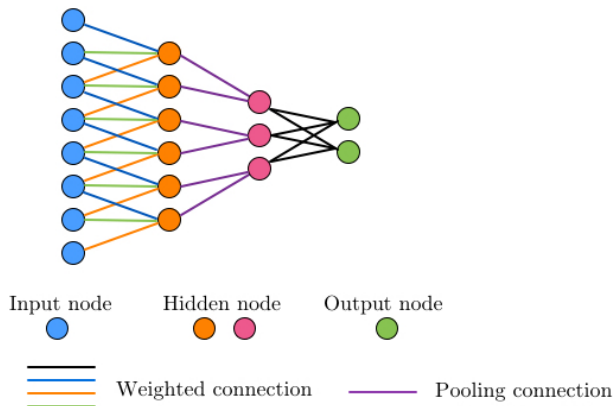


Figure 2: Basic illustration of a node graph of 1D CNN, representation of a typical architecture. The blue, orange and green weighted connections also indicate shared weights within similar colours.

2.3 Transpara

The AI application used in this work is Transpara version 1.4.0 (ScreenPoint Medical, Nijmegen, the Netherlands). The AI application is a tool that assists radiologists with the reading of mammograms [24]. It implements deep learning through training CNNs to find and classify calcifications and soft-tissue lesions. By combining these features, suspicious regions can be identified. The suspicion is graded by values from 0 to 100 (15-95) where 100 indicates the highest level of suspicion (Figure 3). Even if each processed area by the AI system gets a score between 0 and 100, at the end scores below 15 are not shown to the radiologist because they lack relevance. Scores above 95 are capped to 95 because else it

might give the impression that the system is 100% convinced that something is cancer. The whole examination consists either of all four diagnostic images from the standard DM examination, which includes CC and MLO views of each breast, or if DBT was performed during recall, the views from the whole volume are included. Based on the suspicion values of findings in all mammographic examination for one case, another Transpara score is determined such that all examinations fall into different categories from 1 to 10. Transpara scores of 1 equals lowest likelihood of cancer and 10 equals highest likelihood of cancer. The system is calibrated to sort cases into ten groups of roughly equal size, where most of the mammograms with cancerous tissue fall into category 10 (roughly 90 % of the cancers fall into category 10) and are the most likely to have cancer [10]. The remaining mammograms with cancerous tissue are very few and fall into lower categories, obtaining lower Transpara scores. The Transpara score therefore provides an alarming system for suspicious mammograms and reports normal mammograms with confidence [10, 24].

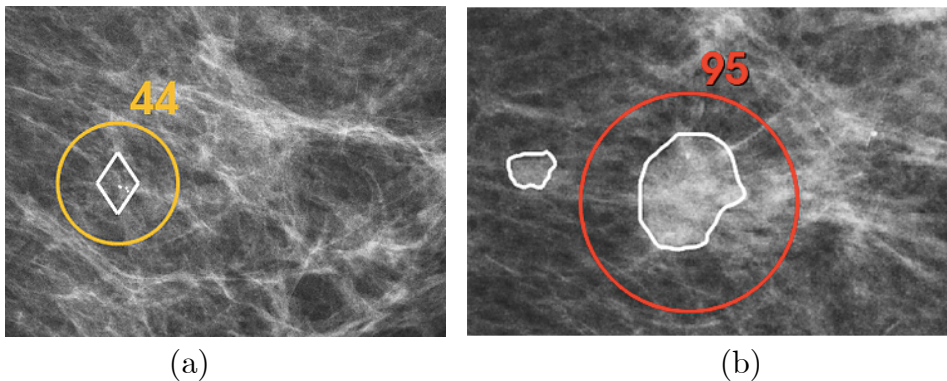


Figure 3: Examples of suspicion values for (a) a calcification and (b) a mass. Image courtesy of ScreenPoint Medical [24].

Like other CNNs based on deep learning, Transpara requires extensive training on images. To work in the desired way, a database of images should provide the AI system with enough material to train, validate and test the system. The preferred database contains mammograms showing abnormalities and healthy mammograms. Transpara has been developed by using a database containing 9000 cases of abnormalities and just as many cases of healthy mammograms [10]. The material in the database originates from different vendors such as Hologic, Siemens, GE Healthcare and Philips Healthcare.

2.4 Mechanical imaging

Mechanical imaging as a medical modality was first described in the late 90's [25]. Mechanical pressure sensors can be used to measure the force or the pressure distribution on the surface of a human body when a load is applied [26]. It has become a tool in the biomedical research field due to that many studies need the use of measuring contact areas and forces. Fuji Film and TekScan are two companies that have developed systems used for these measurements [26]. Both systems are not intended for MI, but pressure sensors from TekScan have been used for MI in studies at the Department of Radiation Physics, Skåne University Hospital, Malmö [12-14, 20, 27, 28].

In breast examinations, the technique of MI makes it possible to find lesions or tumours without using ionizing radiation. The pressure distribution obtained with the mechanical sensors can indicate areas of increased pressure, which is otherwise detected by doctors examining the breast by hand searching for hard nodules. That examination is called palpation. Usually, palpation is an effective method for finding breast and prostate cancer, but mechanical imaging can provide the same or better information in a quantitative and automated fashion. However, all tumours cannot be detected by palpation alone due to different tumour structures, location of the tumour and the experience of the doctor. Through the mechanical information obtained in mechanical imaging, tumours that are not detected by palpation may be detected by mechanical imaging due to more subtle changes in mechanical structures [29, 30]. This was stated by Sarvazyan [29] and have later been supported by Egorov et al. [31] and Dustler [20].

The elastic properties of a material can be described according to Hooke's law:

$$\sigma = E \cdot \epsilon$$

where E is the material constant of Young's modulus, also called the elastic modulus, which is a property quantifying the stiffness of the material. The equation describes the relationship between mechanical stress, σ , and mechanical strain ϵ [32]. Compressive stress is here called pressure, and it has a negative sign since the breast is subjected to a uniaxial load or force, F , in the axial direction on an area, A . The mechanical strain is dimensionless and describes the relative deformation of an object subjected to a stress (Figure 4), as the ratio between the deformed length ΔL and the original length L :

$$\sigma = -\frac{F}{A}, \quad \epsilon = -\frac{L - L_{\Delta}}{L} = -\frac{\Delta L}{L}$$

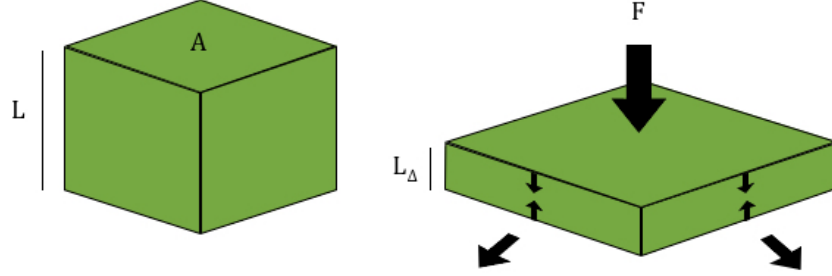


Figure 4: Illustration of deformation of an elastic cubic shape. As the force F is applied to an area A the original length L turns into the deformed length L_{Δ} .

As the breast is compressed during mammography, the applied compression force gives rise to surface stresses. This stress is measured with pressure sensors, and as the deformation (strain) is constant, the measured stress is proportional to the elastic modulus of the underlying tissue. When a material consisting of various components is subjected to a certain force the stresses varies due to the different elastic moduli [20]. Breasts consist of adipose and fibroglandular tissue, which means that the elastic modulus differs between the two tissues. The material modelling is based on the assumption that the material can be considered only experiencing stresses parallel to the applied force and free to deform orthogonally to the direction of the stress. The entire measurement is then the mean stress over a number of ‘cylinders’, where all the cylinders form the entire breast volume which is affected by surface stress. For more details on the formulation of the material model refer to [20]. The positioning of the pressure sensors on the mammographic compression paddle is seen in Figure 5. Egorov et al. have shown that MI can differentiate between benign and malignant tumours since malignant lesions express increased stiffness compared to benign lesions [31].

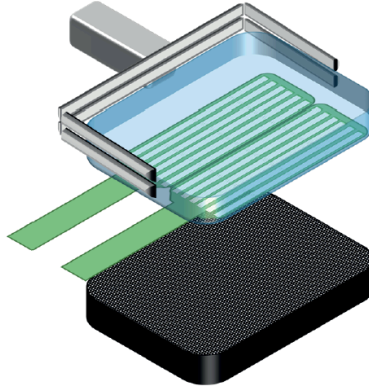


Figure 5: Pressure sensor positioning on the mammography compression paddle [20].

Mechanical imaging itself is a non-ionizing technique as the pressure measurement does not require ionizing radiation. However, the sensors that are commercially produced are not radio-translucent. Since the pressure measurement is performed during another image acquisition, and not the diagnostic image acquisition, the mechanical image must be matched to a low-dose image (pre-exposure image) such that the two images can be compared. The two images are required due to possible difference in breast position, compression force and internal changes of the breast.

3 Materials and methods

3.1 Data collection

The mammographic images and the mechanical images were obtained from a previous prospective study by Dustler et al. [13]. The study used pressure distributions to evaluate the distribution patterns in the screening mammograms, and if any reduced compression force affected the mammographic examination. The authors concluded that the compression plates used did not provide the optimal compression for most women, while it was not caused by a lack of compression force. The mammograms were collected from recalled screening examinations, which for every patient included DM and MI on a Siemens MAMMOMAT Inspiration system (Siemens Medical Systems, Erlangen, Germany).

For MI, the TekScan Iscan 9801 (TekScan, Boston, MA) pressure sensor was used. The pressure sensor elements are oriented as 6 columns with 16 elements in each column (Figure 6). The spatial resolution of the pressure sensor is 12.7 mm. Each column is divided into four parts, each part having four elements that are connected to one common circuit. Two pressure sensors were used for each measurement to cover the whole image detector (Figure 5), resulting in 192 sensor elements.

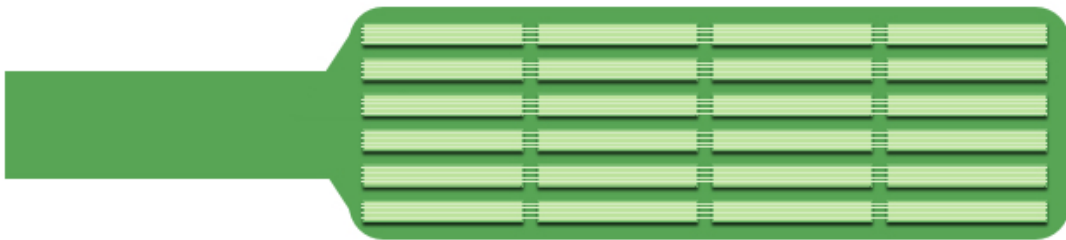


Figure 6: Basic illustration of one TekScan Iscan 9801 pressure sensor.

The AI data was obtained by sending screening images to the software Transpara. The mechanical images were created by writing a script in MATLAB release 2018b (The MathWorks, Inc., Natick, MA), where the pressure matrices for every breast were merged with the low-dose x-ray image. The mean pressure within the breast volume and the mean pressure over the Transpara finding were calculated using Microsoft Excel.

3.2 Suspicious lesion analysis

Screening images of 118 women were analysed in the artificial intelligence software Transpara. Before analysing the images, the name of each image was saved in a specific list of images. The images were put in a folder ‘Send to Transpara’ and then a .bat file started the processing of images through Transpara. It takes some time for Transpara to process the images, usually around a few hours with this number of images (472 diagnostic images). When the images were processed, the results were saved in an Excel file. The Excel file contains all data from all images run in Transpara for those cases where findings were present.

The features found by Transpara were in this work separated into three categories; masses, calcifications, and soft-tissue lesions. Each feature was presented with a certain colour, indicating the type of feature. Red findings represent masses found by Transpara, yellow findings represent calcifications found by Transpara and blue findings represent soft-tissue lesions found by Transpara. Examples of findings are presented in Figure 7.

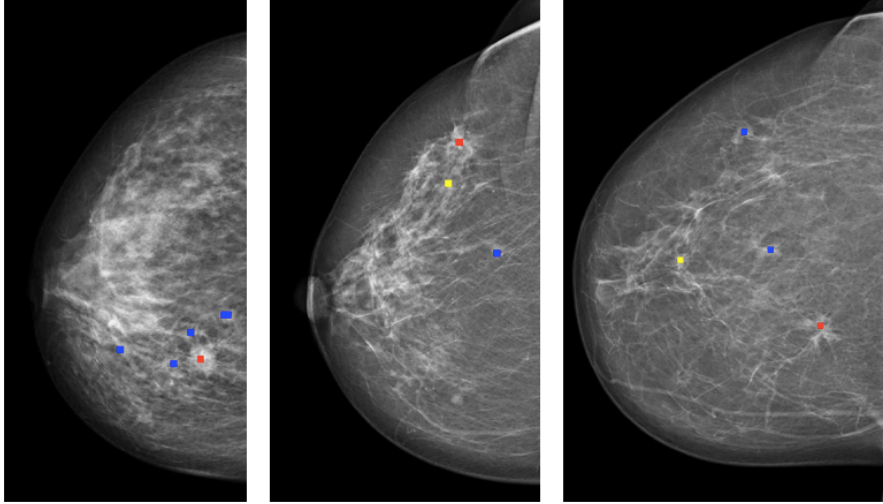


Figure 7: Examples of Transpara findings found in three different cases, showing masses (red), soft-tissue lesions (blue) and calcifications (yellow).

For cases with multiple findings, the finding with the highest score was chosen to be analysed. For those cases where the finding with the highest score was close to the juxtathoracic part of the breast or close to the edges, the next highest finding was chosen instead.

The mechanical images were created by writing a script in MATLAB release 2018b (The MathWorks, Inc., Natick, MA), where the pressure matrices for every breast were merged with the low-dose x-ray image. The mean pressure within the breast volume and the mean pressure over the Transpara finding were calculated using Microsoft Excel.

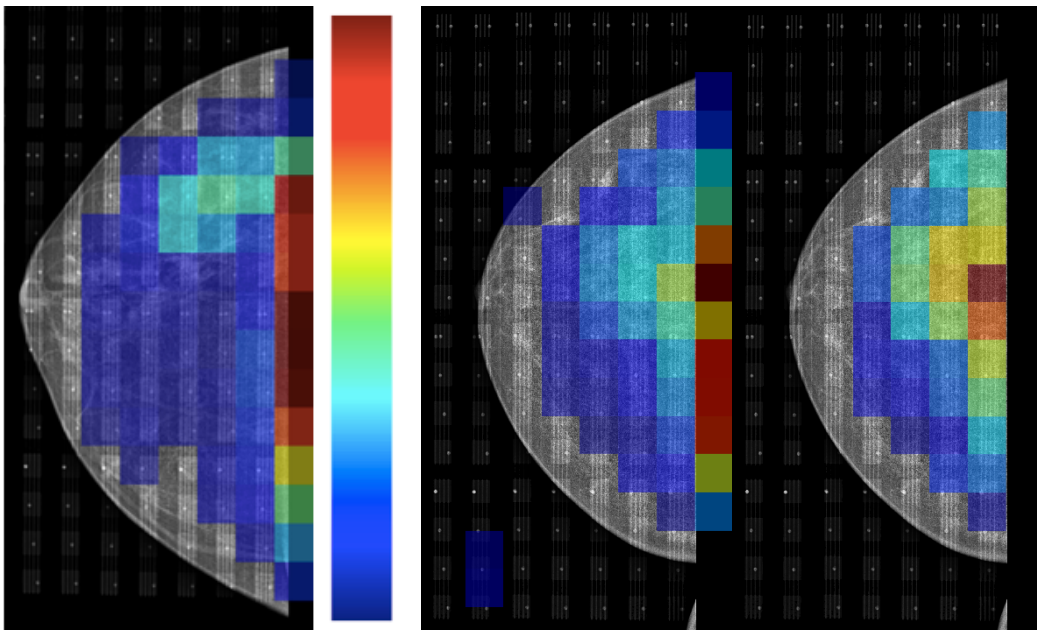
3.3 Mechanical imaging analysis

The mechanical images were acquired and processed by writing a script in MATLAB, release 2018b (The MathWorks, Inc., Natick, MA), where functions already used in a previous study by Dustler et al. was used [13]. The mechanical image was overlaid on the low-dose mammogram of each breast for every woman. The mechanical pressure data was collected as a .asf file and the pressure was extracted through a function that returns the pressure values in a 16x6 matrix (one for each pressure sensor). Since two pressure sensors were used, the two matrices were added together and formed the entire pressure matrix of 16x12

elements. The script merged the new image of the pressure distribution of the breast with the low-dose mammogram.

The mechanical image is a pattern of squares with different colours, representing the relative pressure. The colour scale of the relative pressure ranges from dark blue (low relative pressure) to dark red (high relative pressure). As the pressure sensors are not radio-translucent they are visible in the low-dose mammogram. Visualization of the merged low-dose x-ray image and the mechanical image with a colour scale is presented in Figure 8(a).

The pressure distribution also showed that a higher pressure was subjected to the juxtathoracic part of the breast, seen in Figure 8(a). To visualize the relative pressure distribution within the breast volume, the most inner row and any other pressure value outside the breast volume was removed, see Figure 8(b).



(a)

(b)

Figure 8: (a) The merged image of the low-dose mammogram and the pressure distribution with a colour scale ranging from dark blue (low relative pressure) to dark red (high relative pressure). (b) Transformation of the 'raw' pressure distribution data to the modified pressure distribution, which represents the pressure distribution within the breast volume instead of all pressure data in the whole sensor matrix.

The mean pressure within the breast volume was then calculated. The mean pressure over a Transpara finding was calculated by taking the mean value of the nine pixels that represents the location of the finding. The mean pressure over a Transpara finding was then compared to the mean pressure of the breast, by calculation of the relative increase in pressure at the location of the finding. The method for choosing an area for each measurement of the mean pressure within the breast volume and over a Transpara finding is visualized and described in Figure 9.

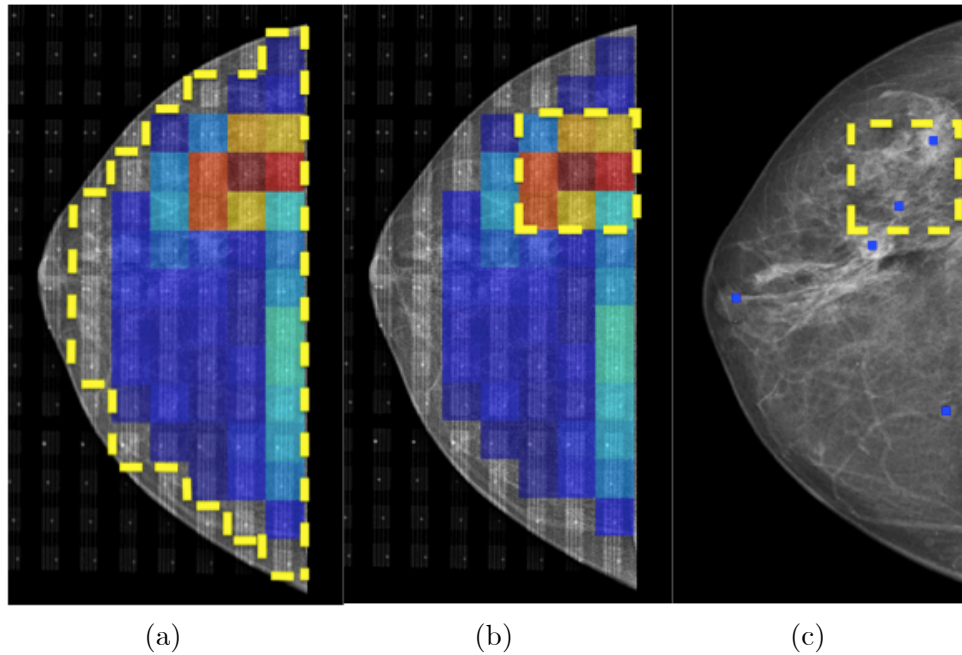


Figure 9: The pressure distribution matrix elements used in the calculation of (a) the mean pressure within the breast volume and (b) the mean pressure over a Transpara finding for a single case. The corresponding region of interest of nine elements from (b) is showed in the diagnostic image with the findings in (c).

3.4 Statistical analysis

To evaluate if there is a significant difference between the increased pressure over a Transpara finding within the three groups (cancer, benign, non-specific) statistical analysis was performed using SPSS version 25.0 (IBM Corp, Armonk, NY). The three groups were statistically analysed to answer the first aim in this study: ‘Relate masses indicated by the AI application to pressure changes relative

to background pressure'. The data of relative increased pressure and the two setups of Transpara scores were tested individually to determine if the residuals were normally distributed. This was tested by performing Shapiro-Wilks test with the null-hypothesis '*The data of increased pressure, total Transpara score and individual Transpara score are normally distributed*' with $\alpha=0.05$ for each group. Based on the result from the Shapiro-Wilks test further tests were carried out with a suitable test for analysis of variance.

Furthermore, the correlation coefficient between the increased pressure and the Transpara scores was tested for each group of cancer, benign and non-specific. This was tested by performing a correlation analysis and to answer the second aim of this study; 'Determine what role the relation plays in clinical mammography'. The non-parametric Spearman's rank order correlation was measured using SPSS, since Spearman's coefficient is appropriate for continuous variables and can describe the relationship as a monotonic function instead of a linear as in Pearson's correlation coefficient.

4 Results

Of the 118 recalled women, findings from Transpara were present for all 118 women in at least one view. Twenty-six women were excluded because the findings were not present in the same view as the pressure distribution view or the pressure distribution was not available.

Of the 92 remaining women, 59 women had findings in both breasts, resulting in a total number of 152 CC images with Transpara findings and pressure distributions. Nineteen masses were found by Transpara in the CC view.

Of the 92 remaining women, 15 had confirmed breast cancer, 35 had benign findings (confirmed by biopsy) and 42 were classified as non-specific. The 42 non-specific cases were not biopsied and diagnosed as healthy. A summary of all cases and their diagnoses are presented in Table 1.

Table 1: The number of cases for each diagnosis, evaluated by a radiologist. One of the women had both IDC and ILC and is therefore listed in both rows.

Diagnose	Number of cases (women)
Cancer (total)	(15)
DCIS	2
IDC	7
ILC	4
Other	3
Benign	35
Non-spec	42
Total number of cases	92

Among the cases with confirmed cancer, masses were found in Transpara in eight of them. The other eleven masses found by Transpara were discovered among women with benign findings (4) and non-specific (7).

The pressure distribution and Transpara findings of a single case are presented in Figure 10. Despite the increased pressure indicating malignancy the woman was diagnosed with a benign finding which was not shown in any of the images. Transpara findings in the right image got scores between 27-37.

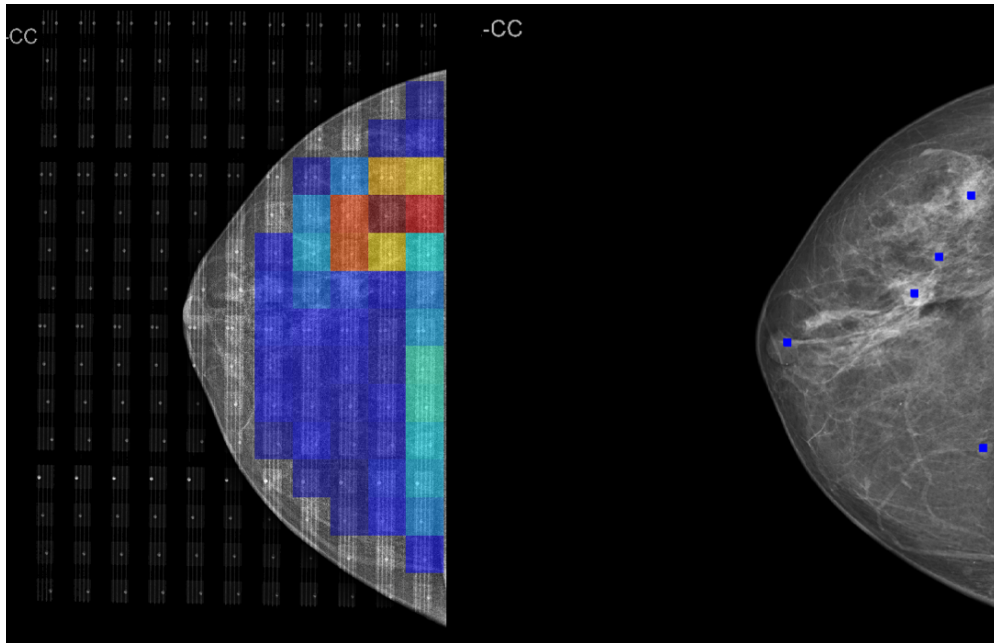
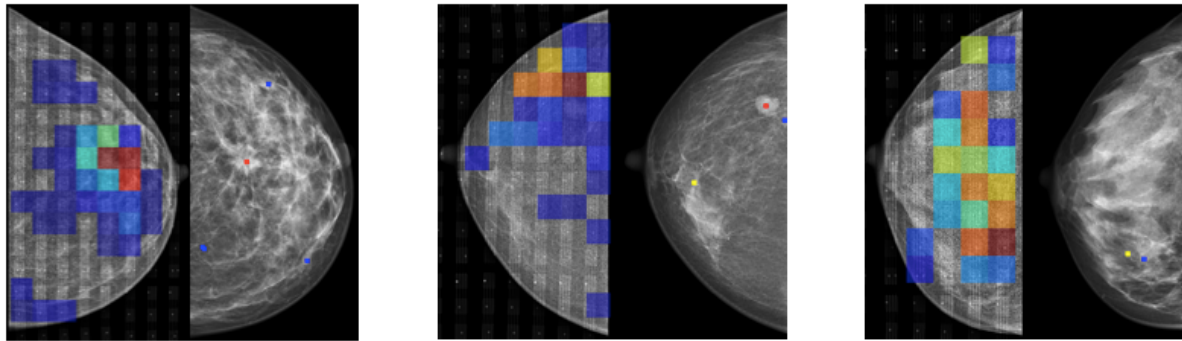
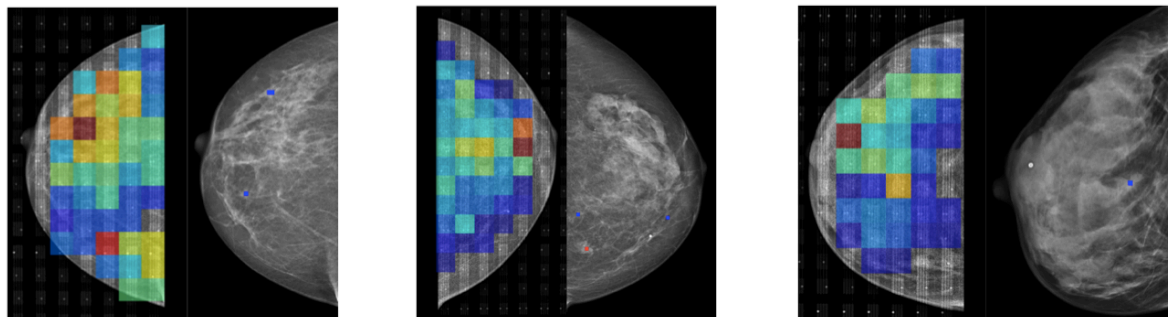


Figure 10: The pressure distribution and Transpara findings of a single case. The most inner row and any other pressure value outside the breast volume has been removed, such that the mean pressure within the breast can be calculated.

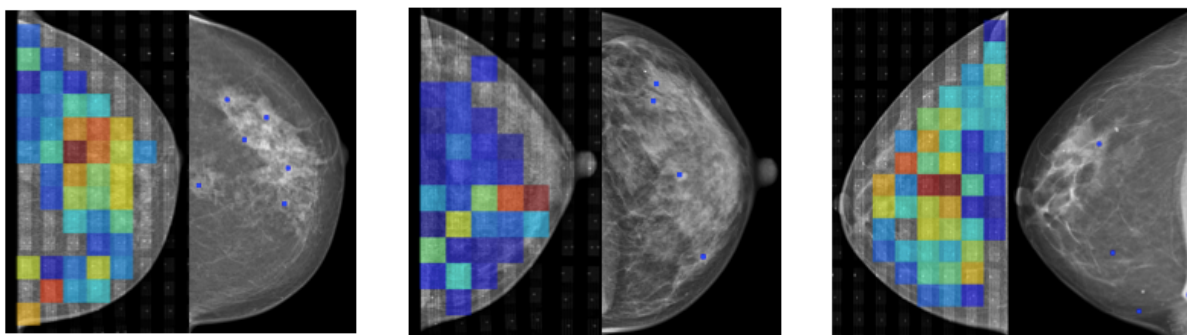
The mean pressure over the breast and the mean pressure over a Transpara finding were calculated and saved in an Excel file. The pressure increase over a finding was calculated for all three groups. Of the 15 women with confirmed cancer, only 12 had a sufficient pressure distribution over the AI-finding or the whole breast. Among women with benign findings, 15 cases could be used. Of the non-specific cases, 19 had a sufficient pressure distribution over the Transpara findings or the whole breast. Example images from the group diagnosed with cancer are presented in Figure 11, of benign findings in Figure 12, non-specific in Figure 13.



(a) (b) (c)
Figure 11: Pressure distribution and Transpara findings of various cases, all diagnosed with breast cancer.



(a) (b) (c)
Figure 12: Pressure distribution and Transpara findings of various cases, all diagnosed with benign abnormalities.



(a) (b) (c)
Figure 13: Pressure distribution and Transpara findings of various cases, all diagnosed as healthy.

A histogram of the pressure difference for all 46 cases is presented in Figure 14, where the cases are 46 women with each one CC image analysed. The histogram shows that for most of the cases, an increase in pressure is present at the location of a Transpara finding, and for some there is actually a decrease in pressure at the location of a finding.

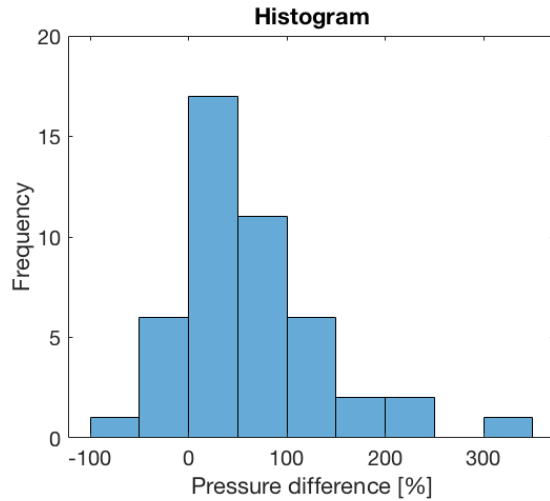


Figure 14: Histogram of the pressure difference for all 46 cases used in the statistical analysis.

The 46 cases analysed were categorized by Transpara, where each mammographic examination got the Transpara Score that describes the likelihood of cancer (1-10). The distribution of Transpara Scores among all cases analysed are presented in figure 15.

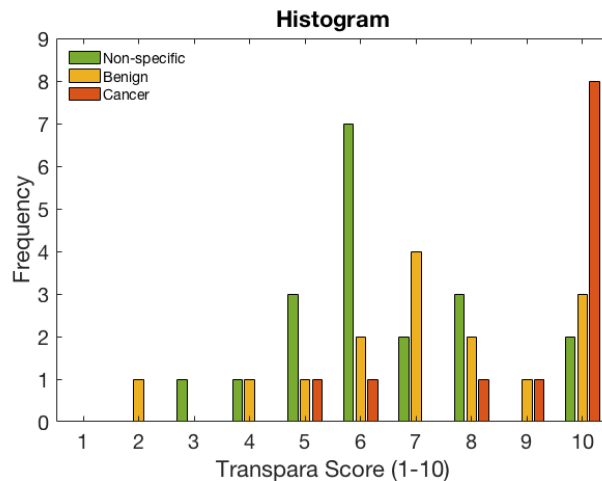


Figure 15: Histogram of the Transpara Scores for all 46 cases divided into the three groups of cancer, benign and non-specific.

For the statistical analysis, data from these three groups of 12, 15 and 19 women respectively were acquired. The Shapiro Wilks test rejected the null hypothesis for all three variables as seen in Table 2, meaning that the data were probably not normally distributed in any case. The variable ‘Score’ represents the Transpara scores of the whole examination (1-10), the variable ‘Score2’ represents the Transpara scores of each finding (0-100) in which the increased pressure is calculated for and the variable ‘Increase’ represents the increased pressure over one finding.

Table 2: Results of the test of normality among the three variables ‘Pressure diff.’, ‘Total Score’ and ‘Ind. Score’. As all p-values were below the significance level, the variables were probably not normally distributed.

	Pressure diff.	Total Score (1-10)	Ind. Score (1-100)
Statistic	0.911	0.925	0.733
df	46	46	46
p	0.002*	0.006*	0.000*

**significant at the 0.05 level*

Since the data was continuous, unpaired and based on the Shapiro-Wilks test also not normally distributed, a non-parametric test with no distributional assumption was used for the analysis of variance. The Kruskal-Wallis rank test was performed with $\alpha=0.05$ to evaluate if the mean ranks of the data were significantly different between the groups. The null-hypothesis used was ‘*The mean ranks of the groups are the same*’ which means that the mean ranks for each variable ‘Pressure diff.’, ‘Total Score’ and ‘Ind. Score’ were compared among the three groups of cancer, benign and non-specific. The results of the Kruskal-Wallis test are presented in Table 3 and as box-plots in Figure 16. The test showed that there was a significant difference in mean ranks between the variables ‘Total Score’ and ‘Ind. Score’, meaning that the mean ranks of both types of Transpara scores differed between groups. The difference in the mean pressure difference between the groups was not significant on the 0.05 level ($p=0.076$). However, the results indicate that the

increase in pressure is higher for the group with cancer as it is significant on the 0.10 level.

Table 3: Results of the Kruskal-Wallis test for the variables ‘Pressure diff.’, ‘Total Score’ and ‘Ind. Score’.

	Pressure diff.	Total Score (1-10)	Ind. Score (1-100)
Kruskal-Wallis H	5.145	10.587	7.443
df	2	2	2
p	0.076	0.005*	0.024*

**significant at the 0.05 level*

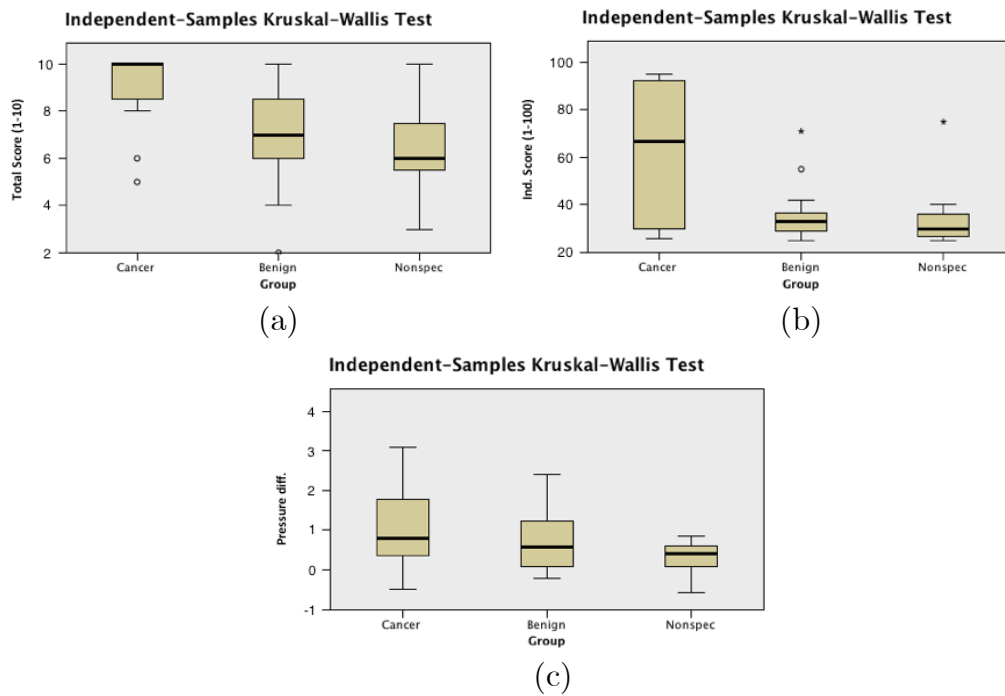


Figure 16: Results of the Kruskal-Wallis test. The null hypothesis was rejected for the variables (a) ‘Total Score’ and (b) ‘Ind. Score’, but remained for the variable (c) ‘Pressure diff.’.

Results of the Spearman's rank correlation between increased pressure and individual Transpara scores in each group, presented in Table 4, showed that the correlation was significant only in one case at the 0.05 level.

Table 4: Results of the Spearman's rank order correlation test between increased pressure and individual Transpara score for the three groups of cancer, benign and non-specific abnormalities.

	Cancer	Benign	Non-specific
N	12	15	19
Spearman's rho	0.592	-0.142	0.242
p	0.043*	0.614	0.317

**correlation is significant at the 0.05 level (2-tailed)*

The Spearman's rho correlation coefficient is 0.592 and significant at the 0.05 level for the group diagnosed with cancer, indicating that the two variables of pressure difference and Transpara score are correlated. For the groups diagnosed with benign and non-specific abnormalities the correlation coefficients are -0.142 and 0.242 but both correlations are not significant at the 0.05 level.

5 Discussion

MI combined with AI can potentially be used to improve the performance of mammography screening, but further evaluation is needed. It has here been shown that an increase in pressure over a finding found by the AI application, tends to be higher for cases with breast cancer. The correlation between increased pressure and Transpara score was found to be strong and significant only for the group diagnosed with cancer. This result suggests that AI and MI find independent markers for breast cancer. How this relationship can be used in clinical mammography remains to be fully investigated.

One of the main limitations of this study is the positioning of the breast between image acquisitions. The mammogram analysed with Transpara is acquired about two weeks before the pressure map is created when the woman is recalled, which makes a comparison between the two modalities difficult. The breast may be positioned in a quite different way between the image acquisitions for example because different radiographers met the woman during screening and recall. The reposition of the breast between the diagnostic image acquisition and the pressure measurement contributes to another major source of error. The pressure sensors are currently mobile and tend to not always be in the same position for every woman. That means that for one woman, the pressure sensors might be slightly tilted in one direction which then is difficult to correct for during the merge in MATLAB. This can also be a source of error in the evaluation of overlap. Ideally, the pressure sensors would be radio-translucent and built-in in the mammography unit.

As seen in Figure 9, the Transpara score of interest is not exactly in the middle of the region of interest. This is often the case, mainly because of the spatial resolution of the pressure distribution data, but also because of the difference in breast positioning during the two image acquisitions. Another impact is how the pressure is distributed in the z-direction, since it is not certain that it will be right above the finding.

Not all AI-findings analysed have the highest score for each case, as earlier mentioned in section 3.2. This is because of lack of pressure data in those regions of the location of the finding or because the finding is located close to the

juxtathoracic portion of the breast. Since the highest individual score is not used in some cases it also affects the analysis by providing an evaluation of the relative increased pressure over an AI-finding which then is not the most suspicious finding of that case.

The results presented in Figure 14 show that most of the 46 women included in the analysis have an increase in pressure over an AI-finding in the breast. These results can sometimes be misleading because of various reasons. In many cases, pressure data is not present throughout the whole breast volume as visualized in Figure 11, Figure 12 and Figure 13, resulting in zero-pressure values in the pressure images. This affects the calculation of mean pressure over the breast volume as well as the calculation of the increase in pressure.

Trends can be seen among the different groups of women, such as that the increase in pressure is present among all cancer cases except 8.3 %. Among the benign cases, 13.3 % have a decrease in pressure over the finding and among the non-specific cases 21.1 % have a decrease in pressure over the AI-finding. Although the cases with the highest increase in pressure are mostly among the cases with confirmed cancer, the statistical analysis did not show a significant difference in ranks of the increase in pressure among the groups ($p=0.076$) on the 0.05 level (Table 2). This study is on relatively low sample size and it is therefore difficult to state anything statistically based on groups of $N=12$, $N=15$ and $N=19$. A previous similar study has shown that there is a difference in increased pressure between cases of cancer and cases of benign lesions [13]. However, the findings in that study were found by radiologists, not by an AI application. As the performance of the AI application Transpara is comparable to the radiologists, the increased pressure is expected to be different between the groups.

The Kruskal-Wallis test was also used to test the difference between mean ranks among groups for the variables 'Total Score' and 'Ind. Score' (Figure 16, Table 3). The variable 'Total Score', representing the Transpara score on the whole examination of the woman, was significantly different between the three groups ($p=0.005$). The variable 'Ind. Score', representing the score in the specific finding analysed, was also significantly different between the three groups ($p=0.024$). This result is expected as a previous study evaluated the performance of Transpara compared to the performance of radiologists, and concluded that it has a detection accuracy compared to the radiologists [11].

The Spearman's rho correlation coefficients for the three groups showed that there was a strong correlation between the increased pressure and the Transpara score for one group, namely the cases with cancer. The correlation coefficient of increased pressure and Transpara score is expected to be weak because the MI and AI are investigating two different aspects of breasts. The AI system is working basically as a radiologist looking for suspicious findings whereas MI is investigating the mechanical properties of the surface of the breast. Since it does not correlate with other AI-findings (present in healthy cases) this result indicates that both methods find independent markers that coincide in cancer. However, it would also be interesting to examine whether the AI-findings among all groups evaluated here correspond to the same findings as the radiologists based on their diagnosis.

For some cases, Transpara classified findings as masses (often resulting in a total Transpara score of 10) but the pressure difference was not noticeably large, seen in Figure 12(b). The diagnosis was biopsy-proven benign, which indicates that the combination of the two techniques could be useful, and maybe has the potential of lowering the recall and biopsy rate.

Measurements of the mechanical stress patterns is also an approximation, since small, multiple cylinders form the breast volume. The stress is calculated as the average pressure over the whole individual cylinder, a potential stiffness within one cylinder increase the average stress over the actual cylinder. It would be interesting for some individual cases to make additional tests or finite element calculations, to investigate whether the approximation is sufficient.

5.1 Future prospects

To make the method easier and more useful on a larger scale, a new radio-translucent pressure sensor should be manufactured. It would also be preferable for it to be built into the mammography equipment. Then the MI can be performed at the same time as the image acquisition of the mammogram. The benefit is then that the positioning of the breast is the same between the two modalities, and AI findings from the mammogram are easier compared with areas of different pressure compared to surrounding breast pressure on the MI. Another benefit is that the women receive a lower radiation dose if both modalities can be

used at the same time, as there is no need for a low-dose image at the same time as the MI.

It would be interesting to study the relation of MI and AI with the new pressure sensor TekScan BPMS 5350 which is used in the clinic today, as it has a better spatial resolution (10.0 mm) and it is more flexible than the TekScan Iscan 9801 pressure sensors. Future studies regarding MI and AI together should investigate the aims of this study with the new pressure sensors.

6 Conclusions

Although it seems like the increased pressure is slightly higher for those cases in the group diagnosed with cancer, the difference is not statistically significant in this study. A material providing larger sample sizes would be needed to further evaluate whether there is a difference.

The correlation found between increased pressure and Transpara score suggests that both MI and AI find independent risk markers that coincide in cancer, and could therefore probably lower the recall rate in mammography screening.

7 Acknowledgements

First I would like to thank my supervisors. Magnus Dustler, for initiating this project and for giving me this opportunity to study MI and AI in breast imaging. I appreciate all your help and guidance during this work. Anders Tingberg, for suggestions, guidance and support.

I would also like to thank the LUCI group for your support and for letting me be a part of your group.

Furthermore, I would like to thank Alejandro Rodriguez-Ruiz (ScreenPoint Medical) for providing me with information and answering questions regarding Transpara.

I would also like to express my gratitude to everyone at the Department of Radiation Physics, SUS, Malmö, who has been welcoming and always providing interesting discussions during lunchtime.

Finally, I would like to thank my family and friends, and a special thank to my boyfriend, August Bjerken, for all the support during these past months.

8 References

1. Cancerfonden. *Bröstcancer*. [cited 2019 January 31]; Available from: <https://www.cancerfonden.se/om-cancer/brostcancer>.
2. Cancerfonden. *Mammografi*. [cited 2019 January 31]; Available from: <https://www.cancerfonden.se/om-cancer/mammografi>.
3. Gur, D., et al., *Digital breast tomosynthesis: observer performance study*. AJR Am J Roentgenol, 2009. **193**(2): p. 586-91.
4. Niklason, L.T., L.E. Niklason, and D.B. Kopans, *Tomosynthesis system for breast imaging*, February 16, 1999, the General Hospital Corp
5. Holt, S., et al., *Digital Breast Tomosynthesis (DBT). A review of the first 1000 patients imaged*. European Journal of Surgical Oncology (ESJO), 2011. **37**(5): p. S26.
6. Helvie, M.A., *Digital mammography imaging: breast tomosynthesis and advanced applications*. Radiol Clin North Am, 2010. **48**(5): p. 917-29.
7. Bahl, M., *Detecting Breast Cancers with Mammography: Will AI Succeed Where Traditional CAD Failed?* Radiology, 2019. **290**(2): p. 315-316.
8. Lehman, C.D., et al., *Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection*. JAMA Intern Med, 2015. **175**(11): p. 1828-37.
9. LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning*. Nature, 2015. **521**(7553): p. 436-44.
10. Rodriguez-Ruiz, A., et al., *Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System*. Radiology, 2019. **290**(2): p. 305-314.
11. Rodriguez-Ruiz, A., et al., *Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists*. J Natl Cancer Inst, 2019.
12. Dustler, M., et al., *Breast compression in mammography: pressure distribution patterns*. Acta Radiol, 2012. **53**(9): p. 973-80.
13. Dustler, M., et al., *Can mechanical imaging increase the specificity of mammography screening?* European radiology, 2017. **27**(8): p. 3217--3225.
14. Fornvik, D., et al., *No evidence for shedding of circulating tumor cells to the peripheral venous blood as a result of mammographic breast compression*. Breast Cancer Res Treat, 2013. **141**(2): p. 187-95.
15. Socialstyrelsen. *Bröstcancer, screening med mammografi*. [cited 2019 January 31]; Available from: <http://www.socialstyrelsen.se/riktlinjer/nationellascreeningprogram/brostcancer-screeningmedmammog>.
16. Andersson, I., et al., *Number of projections in mammography: influence on detection of breast disease*. AJR Am J Roentgenol, 1978. **130**(2): p. 349-51.

17. Cancerfonden and Socialstyrelsen. *Cancer i siffror 2018*. 2018 [cited 2019 May 20]; Available from: https://static-files.cancerfonden.se/Cancer_i_siffror_2018_laddaner.pdf.
18. Poulos, A., et al., *Breast compression in mammography: how much is enough?* Australas Radiol, 2003. **47**(2): p. 121-6.
19. Miller, D., V. Livingstone, and P. Herbison, *Interventions for relieving the pain and discomfort of screening mammography*. Cochrane Database Syst Rev, 2008(1): p. CD002942.
20. Dustler, M., *Pressure distribution in mammography. Mechanical imaging and implications for breast compression*, in *Faculty of Medicine*. 2016, Lund University.
21. *The world Health Organization Histological Typing of Breast Tumors--Second Edition. The World Organization*. Am J Clin Pathol, 1982. **78**(6): p. 806-16.
22. Litjens, G., et al., *A survey on deep learning in medical image analysis*. Med Image Anal, 2017. **42**: p. 60-88.
23. Yamashita, R., et al., *Convolutional neural networks: an overview and application in radiology*. Insights Imaging, 2018. **9**(4): p. 611-629.
24. ScreenPointMedical. *Transpara*. [cited 2019 January 31]; Available from: <http://www.screenpoint-medical.com/transpara>.
25. Sarvazyan, A.P. and A.R. Skovoroda, *Method and apparatus for elasticity imaging*. 1996.
26. Bachus, K.N., et al., *Measuring contact area, force, and pressure for bioengineering applications: using Fuji Film and TekScan systems*. Med Eng Phys, 2006. **28**(5): p. 483-8.
27. Dustler, M., et al. *The effect of breast positioning on breast compression in mammography: a pressure distribution perspective*. in *SPIE Medical Imaging*. 2012. SPIE.
28. Förnvik, D., et al. *Pressure distribution in mammography: compression of breasts with malignant tumor masses*. in *SPIE Medical Imaging*. 2013. Lake Buena Vista (Orlando Area), Florida, United States: Proc. SPIE 8668.
29. Sarvazyan, A., *Mechanical imaging: a new technology for medical diagnostics*. Int J Med Inform, 1998. **49**(2): p. 195-216.
30. Sarvazyan, A.P. and V. Egorov, *Mechanical imaging in medical applications*. Conf Proc IEEE Eng Med Biol Soc, 2009. **2009**: p. 1975-8.
31. Egorov, V., et al., *Differentiation of benign and malignant breast lesions by mechanical imaging*. Breast Cancer Res Treat, 2009. **118**(1): p. 67-80.
32. Wihlborg, G., *Hållfastighetslära*. 2006, Avdelningen för Hållfastighetslära, LTH: Lund University.