NTNU
Norwegian University of
Science and Technology

Classification of Breast Cancer Based on DCE MRI

Ingrid Quist Løkken

Master of Science in Physics and Mathematics
Submission date: June 2017
Supervisor: Pål Erik Goa, IFY

Norwegian University of Science and Technology
Department of Physics

Summary

This master thesis is a part of a project at the MR group at the department of physics at NTNU. The aim of MR group is to make a model which is able to classify the subtype of breast tumours. In this master thesis the focus has been on dynamic contrast enhanced images

The first part of this master thesis consist of trying out different methods for motion correcting images. The reason for this was some unsatisfactory results of the motion correction done in the specialisation project during the fall of 2016. Some of the patients had images with remaining motion after the motion correction algorithm had been applied. The motion correction was now done using normalised cross correlation with 3, 6, and 9 degrees of freedom, and then using another similarity function, namely normalised mutual information. The result was that for some patients increasing the number of degrees of freedom from 3 to 6 would slightly improve the results of the motion correction.

The next part was to coregister the images to a reference image. This was done to make the future combination of images from different MRI modalities easier. The reference image chosen was a T2 MRI image, because this is taken for all patients for anatomical reasons. The method used for coregistration was a software called CMTK, the same we used for motion correction. The similarity function used was normalised cross correlation with 3 degrees of freedom.

Then regions of interest was found for different parameters, we chose to try to classify air, breast tissue in general, benign tumours and malignant tumours. ROIs was found for the included patients, and the pixel values implemented in a matrix. Then a SVM classifier was used on the matrix, to see if it was possible to classify different kinds of pixels in a test set. When only including benign tumours, not malignant, the SVM classifier classified 99.7 % of the pixels correctly. When including malignant tumours the correctly classified rate dropped slightly to 97.4%.

Sammendrag

Masteroppgaven er en del av et større prosjekt hos MR gruppen på insitutt for fysikk på NTNU. Målet til MR gruppen er å lage en modell som kan klassifisere subtyper av brystkreft. I denne masteroppgave har fokuset vært på dynamisk kontrast forsterkede bilder.

Den første delen av oppgaven besto av å gå gjennom bevegelseskorreksjonen gjort som en del av et prosjekt høsten 2016. Resultatene i prosjektet var ikke tilfredsstillende for alle pasienter, og derfor ble det for noen av pasientene prøvd andre parametre. I prosjektet fra høsten 2016 ble bevegelseskorreksjonen gjort med normalisert krysskorrelasjon og 3 frihetsgrader. Nå ble også 6 og 9 frihetsgrader inkludert, itillegg ble en annen likhetsfunksjon, normalisert gjensidig informasjon, prøvd ut. Resultatet var at NCC med 6 frihetsgrader ga litt bedre resultater enn NCC med 3 frihetsgrader for noen pasienter.

I den neste delen av prosjektet ble DCE bildene koregistert til et referanse bilde. Det ble gjort fordi da vil det i fremtiden være mulig å inkludere andre bildemodaliteter, som diffusjons bilder. Som referansebilde ble et T2 vektet bilde valgt. T2 er alltid inkludert i bildeseriene av anatomiske grunner, og finnes derfor for all pasienter. Koregistreringen ble gjort med CMTK, det samme programmet som ble brukt til bevegelseskorreksjonen. Likhetsfunksjonen brukt var NCC med 3 frihetsgrader.

Til slutt ble ROIer, regions of interest, lagd for all inkluderte pasienter. Det ble lagd ROIer for luftområder, brystvev generelt, godartede svulster og ondartede svulster. Informasjonen fra ROIene ble inkludert i en matrise. På matrisen brukte vi en SVM klassifikator, for å se om det vil være mulig å lage en modell som kan klassifisere pixeler i et testsett. Når kun godartede svulster var inkludert klarte modellen å klassifisere 99.7 % av pixelene. Når ondartede svulster ble inkludert sank klassifikasjonsraten litt til 97.4%.

Preface

In this master thesis I have checked the motion correction for some of the patients, coregistered images for the included patients, made ROIs for different kinds of tissue/air for the included patients and performed machine learning on the information from the ROIs to try to classify tissue in breast cancer patients. This paper is part of my master thesis in Biophysics and Medical Technology at NTNU as part of the study program Physics and Mathematics. The work was carried out at department of physics in the Spring 2017.

It is part of a larger project at the MR group at department of physics, and will contribute to their objective of making a software which can be used to diagnose breast cancer by the means of machine learning algorithms.

Trondheim, 2017-06-23

(Signature)

Ingrid Quist-Løkken

Acknowledgment

I would like to thank my supervisor Pål Erik Goa for his great help during my writing of this master thesis. In addition I would like to thank the MR group at the department of physics, and specially Igor Vidic.

I.Q.L

Table of Contents

List of Figures

Abbreviations

AIT Arterial Input Function

CMTK Computational Morphometry Toolkit

DCE Dynamic Contrast Enhanced

DoF Degrees of Freedom

EES Extracellular Extravascular Space

ER Enhancement Ratio

FA Flip Angle

LIBSVM Library for Support Vector Machine

MRI Magnetic Resonance Imaging

MSD Mean Square Difference

NCC Normalised Cross Correlation

NMI Normalised Mutual Information

RMS Root Mean Square

ROI Region of Interest

TE Echo Time

TR Repetition Time

SVM Support Vector Machine

Chapter 1

Introduction

1.1 Introduction

Breast cancer is the most often diagnosed cancer in women [2] and a leading cause of death in women. 23% of all cancers diagnosed is breast cancer, which makes it one of the larges cancers overall, after lung cancer, and breast cancer accounts for 14% of all cancer death. This implies that breast cancer is a type of cancer which affects many people. Developing better diagnosing tools will therefore affect a lot of people. Today the main diagnosing tools are mammography followed by a biopsy. The biopsy might be painful for the patients, and in many cases it is not necessary as the tumour might be benign. A diagnostic tool based on magnetic resonance imaging(MRI) will therefor reduce the number of women who needs biopsies. MRI images, from different modalities, can contain both anatomical and functional information. A combination of these images might therefore provide useful information about the kind of tumour. The information from the different MRI modalities can be combined to make a machine learning model, which can be thought how to see the difference between benign and malignant tumours, and maybe even the subtype of malignant tumours. This would have a great impact on how we look at diagnostic tools, and make MRI an important part of any breast cancer diagnosis. In the future biopsies might not even be needed, computers could take over as the main diagnostic tool for cancers as machine learning algorithms are improved and new information from MRI images are discovered.

In this project the main focus has been on reading dynamic contrast enhanced(DCE) MRI images for data processing, and thereafter try to develop a model which correctly classifies

different kinds of tissue in the breast. Some images were motion corrected again, using different parameters than in the specialisation project done during the fall 2016[3]. The number of degrees of freedom(DoF) in the motion correction was increased, to see whether it would yield better results, as we were not satisfied with the motion correction for all patients using normalised cross correlation(NCC) with three DoF. In addition we found documentation during the fall of 2016 [4] that lead us to normalised mutual information(NMI) as a possible similarity function for motion correction. In this thesis NMI therefore was included as a possible method for motion correction.

To be able to compare images and information from different MRI modalities it is important to be sure that they in fact are directly comparable. Since the images are from patients, which might lie in the MRI scanner for a long time, it is likely that the patients moved somewhat during their time in the scanner. Therefore one of the imaging series has to be set as the reference, so all other images can be coregistered to the common reference image. The reference image chosen has to be of a MRI modality which is included for all patient. Since a T2 weighted image always is included for anatomical reasons is might be a good reference image. DCE images can then be coregistered to their T2 reference, and then later it will be possible to compare other MRI modality images with the DCE images, given that all the images are coregistred to the same reference.

In coregistered images it is now possible to collect information from different parts using regions of interest (ROIs). ROIs can be made using different kinds of software, and of areas of your choice. It was therefore decided that ROIs of air, breast tissue and tumour tissue would be made. Breast tissue ROIs consisted of both fibroglandular tissue and fat tissue, as in many patients it was hard to make accurate ROIs for only fibroglandular tissue. The tumour might be benign or malignant. The information from the ROIs can then be combined in tables or matrices and then used to make a classification model.

This project used machine learning algorithms on the matrices. A support vector machine(SVM) classifier was implemented using free software, and the matrices made from ROIs was used to make the model. Then the model was tested using a test set with values either from different ROIs from the same patient, or from ROIs from another patient. The preliminary result from the model made is included. Including more patient ROIs when training the model will probably make the model more stable, and increase the fraction of correctly classified pixels.

Chapter 2

Theory

Sections 2.1 - 2.4 are largely based on the theory from the specialisation project done during the fall of 2016 [3]. The rest of the theory is written specially for this master thesis.

2.1 DCE MRI

In DCE MRI of the breast a contrast agent is used to enhance the contrast between normal tissue and tumour tissue[5][6]. When injected, the contrast agent will follow the blood and leak out into the extracellular matrix. Malignant tumours in the breast show faster uptake of contrast agent than normal tissue and benign tumours[7] [8]. This is assumed to be because the blood vessel wall in malignant tumours are more leaky than normal blood vessel walls. The contrast agent will therefore be able to leak out into the extracellular matrix of the tumour faster than into the extracellular matrix of normal tissue and benign tumours. This gives the enhanced contrast agent uptake in tumours, and specially in malignant tumours.

The accumulation of contrast agent in the extracellular matrix of tumours is utilised in DCE MRI where a paramagnetic contrast agent is used to increase the signal. A paramagnetic contrast agent is often used in T1 weighted MRI imaging as it alter the T1 relaxation time in the tissue at which it accumulates. The reason for this effect comes from the fact that the contrast agent contains metal ions. These metal ions usually have unpaired electrons, and the metal ions in the contrast agent becomes dipoles when put in a magnetic field [9]. These dipoles will make a large fluctuating magnetic field, at least this is how it will feel for the protons nearby because electrons have a 700 times greater magnetic moment than protons. This will then enhance the

proton relaxation, specially if the frequency is close to the Larmour frequency. The more unpaired electrons the paramagnetic metal ions have, the larger the effect on the protons to be enhanced will be.

DCE MRI is a dynamic imaging method, meaning that several images is taken over a course of time. During the acquisition time of approximately seven minutes, eight image series are acquired. Each series consists of 60 slices, covering the whole breast. The first series is taken before the contrast agent is injected, giving a reference image. Then seven series are acquired, with approximately one minute between each acquisition, after the contrast agent has been injected. It is then possible to look at the time development of the signal. If the reference image, the pre-contrast image, is subtracted from the post-contrast images the tumour will light up in the images.

From these images it is possible to collect several parameters. There are two methods for collecting information about the parameters, quantitative and semi-quantitative [10]. Semi-quantitative parameters are parameters which can be calculated directly from time signal intensity curves acquired from the images [1]. Examples include initial slope, signal enhancement ratio, washout, time-to-peak etc. Quantitative parameters on the other hand depend on parameters directly measured during the acquisition, like K^{trans} the transfer constant of diffusive transport of a contrast agent [11].

2.2 Contrast

In most DCE MRIs the contrast agent used is a non-specific agent consisting of small molecular agents which readily distributes in the extracellular matrix [12]. In cancerous tissue the vasculature is abnormal because of the high degree of angiogenesis occurring. Angiogenesis causes blood vessels to grow and tumour cells release factors which encourage angiogenesis to take place [13]. Without a developed blood vessel network the tumour will not be able to grow beyond the size of a few millimetres, because of the diffusion length of oxygen and nutrient in tissue(about 150 μm)[12].

The growth of new blood vessels is fast, because the tumour wants to grow. As a result the new blood vessels in the tumour is leaky and the structure is chaotic. Since the blood vessels are more leaky in the tumour than in the rest of the tissue, the concentration of contrast agent

leaking out into the extracellular matrix will be greater in the tumour [12]. This is utilised in DCE MRI, where a dynamic imaging series is acquired. The tumour will show a faster increase in signal intensity compared to the surrounding normal tissue. Another effect of the leaky blood vessels in the tumour is the fast washout of contrast agent. The washout is fast, as well as the uptake, and during the acquisition there will be a maximum signal about 2 minutes after injection, and then a rapid washout, see figure 2.1.

In the presence of a T1 affecting contrast agent the T1 of the tissue is reduced [11] by the following equation

$$\frac{1}{T_1'} = \frac{1}{T_1} + r_1 C \tag{2.1}$$

T_1' is the new, reduced T_1 value, T_1 is the pre contrast value, r_1 is the relaxivity and C is the concentration of contrast agent. Equation 2.1 can be rewritten using the fact that R_1 is $1/T_1$ and R_1' is $1/T_1'$, giving the relaxation rates,

$$R_1' = R_1 + r_1 C. \tag{2.2}$$

The contrast agent reduce T1 because of its paramagnetic effect. Paramagnetic metals have unpaired electrons in their outer shells, which becomes magnetic dipoles when placed inside a magnetic field. The protons nearby will then experience a large fluctuating magnetic field. If these fluctuations have a frequency close to the Larmor frequency the result is an enhanced proton relaxation [9].

In equation 2.2 a linear relationship between the relaxation rate and contrast agent is assumed. This implies that all protons have the same access to the contrast agent on a time scale which is equal to the T1 relaxation time of the tissue [9]. This is called fast exchange and can be described by the correlation time τ between the intracellular and the extracellular compartments.

$$\frac{1}{\tau} >> \frac{1}{T_{1,1}} - \frac{1}{T_{1,2}} \tag{2.3}$$

$\frac{1}{\tau}$ is the rate of water exchange between compartments with relaxation rates $\frac{1}{T_{1,1}}$ and $\frac{1}{T_{1,2}}$ for the intracellular and extracellular compartments respectively [9]. The resulting signal will depend on the volume fraction of each compartment, and give an effective R1,

$$R_{1,eff} = fR_{1,1} + (1-f)R_{1,2}, \tag{2.4}$$

where $R_{1,eff}$ is the effective relaxation rate, $R_{1,1}$ is the relaxation rate in the first compartment and $R_{1,2}$ is the relaxation rate in the second compartment. f gives the fraction of each volume.

2.2.1 Spoiled Gradient Sequence

To acquire DCE images a spoiled gradient flash 3D sequence is utilised. In this sequence a short echo time (TE), a short repetition time (TR) and a small flip angle is utilised to make a T1-weighted image. Since spoiled gradient sequences are used, which are fast, we can image the contrast agent contamination of the tumour tissue, while contamination of normal tissue will be to slow, and out of the time period we use.

The signal equation for a spoiled gradient echo sequence is given by the following equations [14]:

$$M_{\parallel} = M_0 \frac{(1 - e^{-TR/T_1}) \sin \alpha}{1 - e^{-TR/T_1} \cos \alpha} \tag{2.5}$$

$$M_{\perp} = M_{\parallel} \sin \alpha e^{TE/T_2^*}. \tag{2.6}$$

$M_{\parallel}$ is the transverse magnetisation,$M_{\perp}$ is the longitudinal magnetisation, TR is the repetition time, TE is the echo time, α is the flip angle and T_1 and T_2^* are properties intrinsic to the tissue.

When the tissue to be imaged has been enhanced by a contrast agent the T1 in 2.5 has to be substituted with the $T1'$ found in equation 2.4 by using the fact that $T1 = 1/R1$.

2.3 Semi-Quantitative Model

It is normal to determine a time-signal intensity curve from the ROI around the tumour. From this curve the initial slope, time-to-peak and wash-out can be found. An example curve is showed in figure 2.1

The initial slope can be seen in figure 2.1 as the part of the slope in the early postcontrast phase. A steep initial slope/early enhancement is often correlated with malignant tumours [15][16][1]. In the intermediate and late postcontrast phase a rapid and early wash-out, which can be seen as line III in figure 2.1, is a sign of malignancy. On the other hand a steady, increasing line, line Ia and Ib in figure 2.1, often is a sign of a benign tumour. A plateau shaped curve,

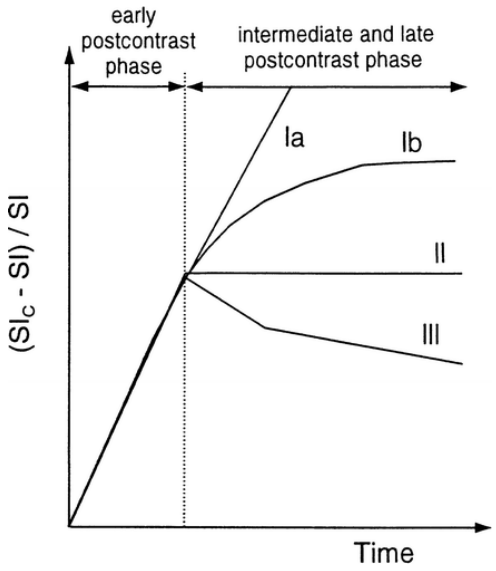


Figure 2.1: The expected time signal intensity curve from a breast tumour. The figure is from an article by Kuhl [1]. Curves Ia and Ib is typical for benign tumours, while curve II and III are typical for malignant tumours.

line II in figure 2.1 is as well a sign of malignancy [1] [16].

The signal enhancement ratio is given by the following equation,

$$ER = \frac{S_c - S_0}{S_0}. \tag{2.7}$$

From this equation it is possible to calculate the enhancement between any of the acquisition series and the reference image, the precontrast image. S_0 in equation 2.7 is the precontrast signal, while S_c is the postcontrast signal in any of the acquisitions.

The slope of the postcontrast phase can be found using the formula for a linear slope as the important information lies in whether it is zero, positive or negative [10]. Equation 2.8 was used for this purpose.

$$Slope = \frac{y_2 - y_1}{x_2 - x_1} \tag{2.8}$$

When using this formula on the DCE MRI data y_2 is the signal at the last time point, y_1 is the signal at maximum if the maximum occurs before the last time point, if not then y_1 is at the second time point after contrast injection. x_2 is the last time point and x_1 is the time point at y_1.

2.4 Quantitative Model

A quantitative or pharmacokinetic model models how the concentration of contrast agent varies between the intravascular and the interstitial space. From this it is possible to find measures of related to tumour blood flow, microvasculature and capillary permeability [16]. The two compartment model, or Tofts model [11], is the mostly used approach to find the parameters mentioned above. This model is used to measure the contrast agent exchange between the extracellular extravascular space (EES) and the blood plasma, see figure 2.2. The plasma concentration is given by:

$$C_p = \frac{C_b}{1 - Hct} \tag{2.9}$$

C_p is the plasma concentration, C_b the blood concentration and Hct is heamatocrit, a factor which relates the two concentrations, typically 42%.

The following equation can then be found from the flow limited model [17][18],

$$\mathbf{v}_e \frac{dC_e(t)}{dt} = K^{trans}(C_p(t) - C_e(t)). \tag{2.10}$$

K^{trans} is the volume transfer constant, v_e is the volume of EES per unit volume of tissue and $C_e(t)$ is the EES concentration [18]. Convolving equation with the impulse response function [17], $K^{trans}exp(-k_{ep}t)$, gives the tissue concentration

$$C_t(t) = \mathbf{v}_p C_p(t) + K^{trans} \int_0^t C_p(\tau) e^{-k_{ep}(t-\tau)} d\tau \tag{2.11}$$

where k_{ep} the flux rate constant between extracellular extravascular space (EES) and plasma. The relationship between K^{trans}, k_{ep} and v_e is given by

$$k_{ep} = \frac{K^{trans}}{v_e}. \tag{2.12}$$

When using a quantitative method in DCE-MRI this relationship is useful as they are considered the three standard kinetic parameters [18]. k_{ep} can be found from the shape of the tracer concentration vs time data. To calculate K^{trans} and v_e on the other hand, you need access to the absolute values of the tracer concentration. These values depend on the balance between capillary permeability and blood flow in the tissue.

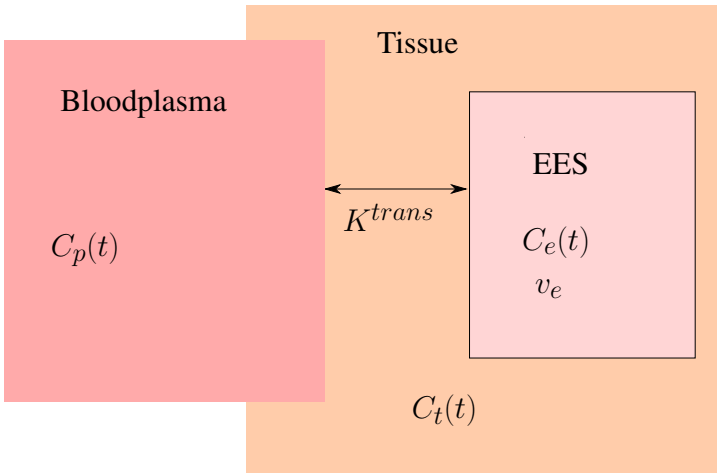


Figure 2.2: The two compartment model. Blood plasma has a concentration $C_p(t)$ of contrast agent. Some of this will be transferred into the tissue and the EES, the amount can be found from the volume transfer constant K^{trans}. Since some contrast agent is transferred into the tissue, the concentration of contrast agent increase, giving a concentration of $C_t(t)$ in tissue and a concentration $C_e(t)$ in the EES. v_e is the volume of EES per unit volume of tissue.

To calculate the extracellular extravascular volume fraction the arterial input function (AIF) has to be measured. As there are no major arteries in the breast a population-averaged AIF has to be used to model the time-dependent contrast concentration in the blood plasma[19]. The concentration of contrast agent in a voxel will be dominated by the contribution from the ESS, but in addition there will be a contribution from the intravascular space.

2.5 Motion Correction and Coregistration

When motion correcting or coregistering MRI images it is important to chose the correct reference image. The reference image is the image used as the correct image. This image is going to be used when correcting all the other images, either for motion, or for coregistering. When motion correcting DCE MRI images it is usual to use the first image in the acquisition series as the reference image, the pre-contrast image. Then all the post-contrast images can be corrected with respect to this image. When coregistering images it is not necessarily the reference image in the same acquisition series which is to be used, it might be an image from another MRI modality, like a T2 weighted image with or without fat-suppression.

When motion correcting or coregistering images a similarity function is used to see how similar the images are to the reference image. There are several possible function. In this project

two types of similarity functions where used, normalised cross correlation and normalised mutual information. Typically root mean square(RMS) is the chosen function when correcting an image based on an other image. This would work well if the images where anatomical, but not when the images are dynamical as in DCE. In the DCE images there will be a contrast uptake which is shown as an increase in intensity in areas affected. RMS uses the change in value(intensity) between pixels in the reference image and in the image to be corrected to calculate the displacement. This will make the algorithm consider the contrast uptake as movement, and thereby make it correct for motion which has not happend.

2.5.1 Normalised Cross Correlation

One similarity function which can be used to coregister or motion correct images is normalised cross correlation (NCC). The equation for NCC is the following [20]

$$NCC = \frac{1}{n} \sum_{x,y} \frac{(f(x,y) - \bar{f})(t(x,y) - \bar{t})}{\sigma_f \sigma_t}, \tag{2.13}$$

where n is the number of pixels, $\bar{f}$ and $\bar{t}$ is the averages of f and t respectively and σ_f and σ_t is the standard deviations of f and t respectively. f(x,y) is here the image to be coregistered or motion corrected and t(x,y) is the reference image.

2.5.2 Normalised Mutual Information

Another similarity function for motion correcting or coregistering images is normalised mutual information. Mutual information as a similarity function was first proposed by Viola [21] and Collignon [22]. They showed that mutual information could be used to align images taken using different modalities well, with accuracy and robustness [23]. The mutual information is given by the following equation [24]

$$MI = I(A,B) = H(A) + H(B) - H(A,B) \tag{2.14}$$

where

$$H(A) = -\sum_a p_A(a) \log p_A(a) \tag{2.15}$$

is the entropy of image A.

$$H(B) = -\sum_{b} p_B(b) \log p_B(b) \tag{2.16}$$

is the entropy of image B. And

$$H(A, B) = -\sum_{a,b} p_A B(a, b) \log p_{A|B}(a|b) \tag{2.17}$$

is their joint entropy.

Normalised mutual information(NMI), as given in [4] , [24] and [25], is given by

$$NMI = \frac{H(A) + H(B)}{H(A, B)}. \tag{2.18}$$

2.5.3 Degrees of Freedom

Another important parameter in motion correction and coregistration is the number of DoF used. Three DoF means only translational parameters are changed. If you choose six DoF then rotation is included as well and with nine DoF the anisotropic scale is included. Anisotropic scaling is when the object in non-uniformly scaled. Twelve DoF includes shear strain as a parameter [26]. These parameter are based on the transformation of object that are rigid bodies. CMTK does not offer to increase the number of DoF further than twelve.

2.5.4 Coregistration

Coregistration is a method to more easily compare images directly. When two images represent the same feature, you want to be sure that the images are directly comparable, so that a specific pixel in one of the images can be found in the other image, and in that way the information from the pixel can be compared. Lets say you have a small tumour in the middle of a DCE image. Then you have diffusion images from the same patient. It would then be nice to look at the diffusion and the contrast uptake in the same area. But to be able to do so, you need to know that it actually is the same region in the body. In MRI this can be hard, as the patient lies in the MRI scanner for a long time, up to an hour. Therefore there might have been movement between the acquisition series, and it is not enough to just motion correct the images within the same series. A solution is therefore to find a common reference image, and then motion

correct, or coregister, the images to this reference. The reference image could be a T1 weighted image, a T2 weighted image, a T2 weighted image with fat suppression, or another MRI image. The advantage of this is e.g. that area in a DCE image and in a diffusion image can be directly compared.

2.6 Machine Learning

The theory about machine learning is based on the course Machine Learning given by Stanford University and professor Andrew Ng, which can be found at coursera.org. The course is based on several books, [27] [28] [29] [30].

Machine learning is the branch of computational science which deals with how computers can learn from experience, [31]. Samuel et al. first proved that it is possible to, with very little programming, to make computers learn from experiences. The first example was a computer programmed to play checkers, which after some playing time could beat the human player which had made the program.

There are two main types of machine learning, supervised learning, which consists of classification problems and regression problems, and unsupervised learning. The two kinds differ, and the method for solving them are different. In unsupervised learning problems the computer is provided with some data and told to find the pattern, if there is one. In classification problems on the other hand, the computer is provided with a training set, and known classes in which it should classify the test data.

2.7 Supervised Learning

In supervised learning problems we have a hypothesis function. To find the optimal hypothesis function we often use a cost function to find the accuracy of our hypothesis function and gradient descent to find the optimised parameters to put into our hypothesis function. Then it is normal to divide the data into a training set and a test set. The advantage of this is that you train the classifier/model using the training set, and then use the test set to test the classifier/model and see how it does on other cases than the cases used to train it. In the optimal case 80 % of the data is in the training set and 20% in the test set, in accordance with the Pareto principle or 80/20 rule [32].

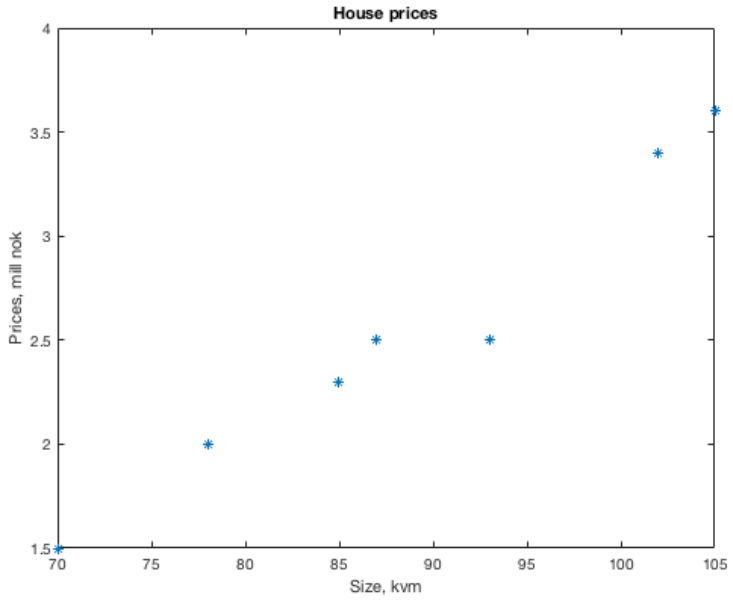


Figure 2.3: House prices as a function of house size in m^2.

2.7.1 Regression Problems

A typical regression problem is to decide the price of something given the prices of similar object. Say you want to decide the price of a house, there are many factors which contributes to the price; the size, number of floors, number of bedrooms etc. But lets say you only know the size of the house. Then you know the prize and the size of some houses in the area. If you then plot the prize as a function of size, see figure 2.3, and then fit a linear line to the data, see figure 2.4, it is possible to estimate the price of the house given the size of the house.

So lets say we have a house which is $97\mathrm{m}^2$, from figure 2.4 we can see that this house will have a price of approximately 3 million NOK. This method of classifying problem is called linear regression. Here the hypothesis function is

$$\hat{y} = h_\theta(x) = \theta_0 + \theta_1 x. \tag{2.19}$$

As you can see this is an equation giving a straight line with parameters $\theta 1$ and $\theta 2$.

Then the cost function can be found from the following equation

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(\hat{y}_i - y_i\right)^2 = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x_i) - y_i\right)^2. \tag{2.20}$$

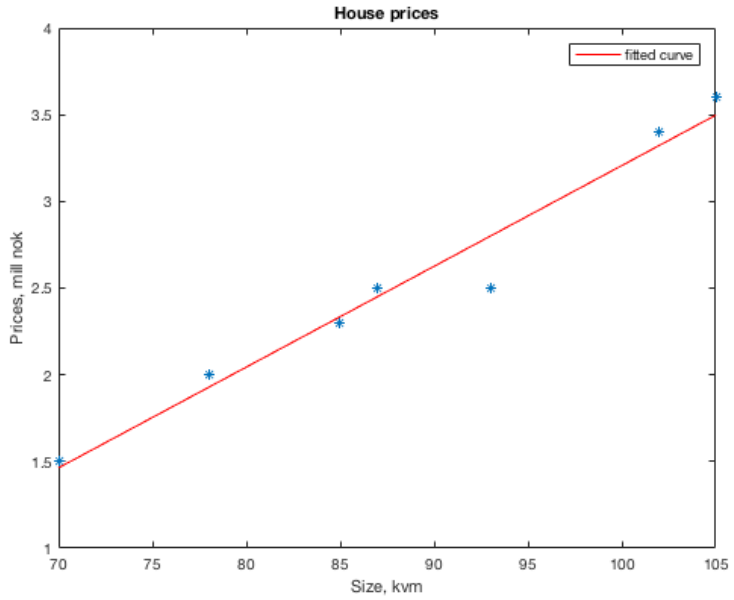


Figure 2.4: The linear regression of the points plotted in figure 2.3.

Where m is the number of samples, $\hat{y}_i = h_\theta(x_i)$ is the predicted value and y_i is the true value.

Now, it is clear that predicting a house price only from the size of the house will in many cases give a wrong estimate. To make a more likely estimate it is possible to add features. This makes the task to find the optimised function more complex. The method used to find the optimised function is often gradient descent. Gradient descent is given by the following equation

$$\theta_j = \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) \cdot x_j^{(i)} \text{ for j = 0:n} \tag{2.21}$$

θ_j is the parameter to be optimised, α is a chosen constant for the step size, which can be regulated, m is still the number of samples, $\hat{y}^i = h_\theta(x^i)$ is the predicted value and y^i is the true value, n is the number of features. Superscript indicates training example, while subscript indicates feature. Gradient descent is used to find the smallest value for θ. It works by iterating over all values, and chooses the lowest value in every round in hope of finding a minimum, see figure 2.5. As seen from the figure it is not certain that the algorithm finds the global minimum. The minimum it finds might only be local minimum, and if this is the case the parameter θ_j might not be set to the lowest possible value.

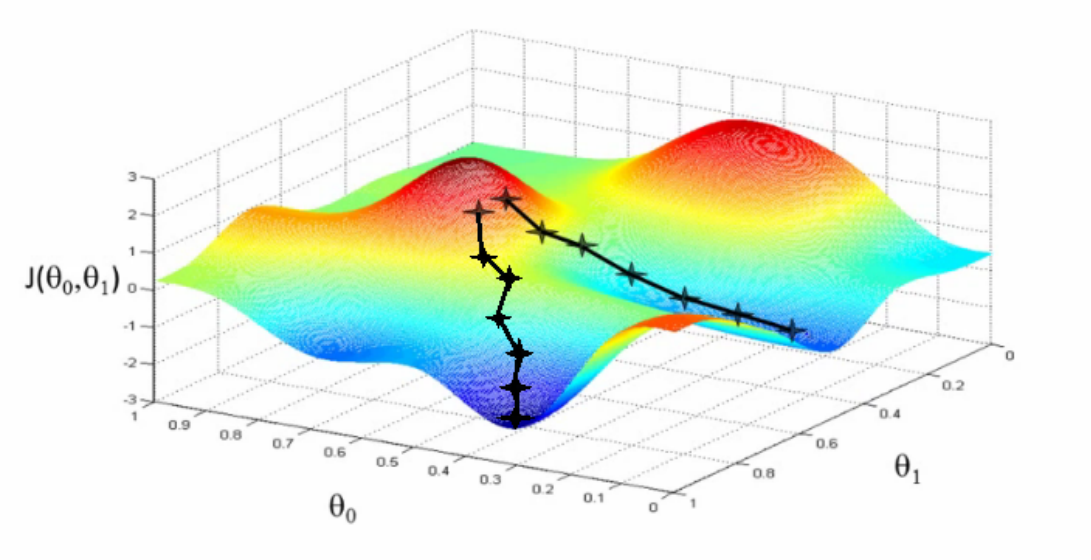


Figure 2.5: Gradient descent. The figure is borrowed from the machine learning course at Stanford University and is made by professor Andrew Ng.

2.7.2 Classification problems

In classification problems we try to classify data into classes. An example could be that you have a house, and you want to predict whether it will sell for more or less than 2 million NOK. The goal in a classification problem is to find the line that gives the best margins. In the easiest example there are only two classes, but it is possible to have several. When classifying into several classes more training is needed.

The easiest algorithm to classify items is linear classification. In linear classification a straight line is drawn, which correctly classifies the whole training set. The optimal line will be the line which is the furthest from both sets. SVM, which is presented in one of the next sections, is a linear classifier.

Logistic Regression

Another kind of classification algorithm is logistic regression. In logistic regression we use the Sigmoid function,

$$g(z) = \frac{1}{1 + e^{-z}}, \tag{2.22}$$

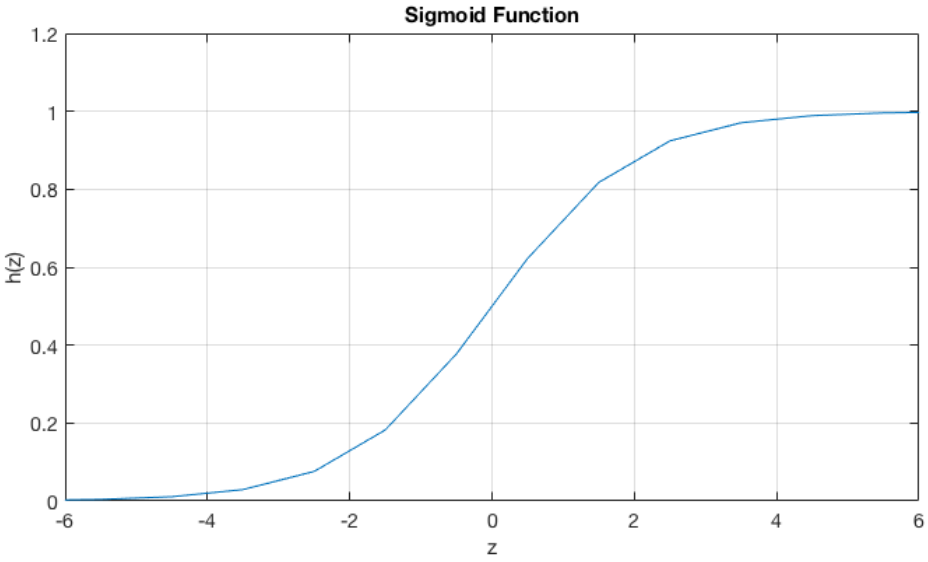


Figure 2.6: The sigmoid function from equation 2.22.

to construct our hypothesis function, see figure 2.6. If we use $z = \theta^t x$ in the sigmoid function, 2.22 we get the hypothesis function

$$h_\theta(x) = g(\theta^t x) = \frac{1}{1 + e^{-\theta^t x}}. \tag{2.23}$$

$h_\theta(x)$ should satisfy the following criteria: $0 \leq h_\theta(x) \leq 1$. Now h_θ, the hypothesis function, gives us a probability. If, lets say $h_\theta = 0.4$, there is a 40% chance for an output of 1, y = 1, and 60 % chance for an output of 0, y = 0. It is normal to use the following decision boundary

$$h_\theta(x) < 0.5 \rightarrow y = 0$$
$$h_\theta(x) \geq 0.5 \rightarrow y = 1.$$

Since $h_\theta(x)$ is equal to $g(\theta^t x)$, this gives that a $z = \theta^t x = 0 \rightarrow g(z) \geq 0.5$

The cost function for logistic regression will be

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)}\log(h_\theta(x^{(i)})) + (1 - y^{(i)})\log(1 - h_\theta(x^{(i)}))]. \tag{2.24}$$

and with $h_\theta(x)$ replaced by the sigmoid function it becomes

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)}\log(\frac{1}{1+e^{-\theta^t x}}) + (1-y^{(i)})\log(1-\frac{1}{1+e^{-\theta^t x}})]. \tag{2.25}$$

The gradient descent equation will then be the same as in 2.21 and the partial derivative will be

$$\frac{\partial}{\partial\theta_j}J(\theta) = \frac{1}{m}\sum_{i=0}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)} \tag{2.26}$$

Regularisation

When classifying data it is important to regularise the data. If not the algorithm might overfit the data, so that it work well on the training set, but poorly on the test set. Regularising the data means adjusting the weight of the variables/features θ_j. This can either be done directly in the hypothesis function, or more easily in the cost function

$$J(\theta) = min_\theta\ \frac{1}{2m}\ \left[\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 + \lambda\ \sum_{j=1}^{n}\theta_j^2\right] \tag{2.27}$$

Here we multiply all θ_j by a constant λ. This makes it possible to change the impact of all the θ_js at the same time. It is important to remember that if the regularisation parameter λ is chosen to large, it might smooth the curve, and the result will be underfitting instead of overfitting.

Support Vector Machine

In SVM we replace the sigmoid function we used as our hypothesis function in logistic regression with a function called Hinge loss [33]. From the cost function in equation 2.25 we alter the first term so that it is 0 when $\Theta^T x$ is greater than 1. When $\Theta^T x$ is less than 1 it will output a decreasing line, see figure 2.7. The second term of the cost function is altered so that it outputs 0 when $\Theta^T x$ is less than -1, and an increasing line when $\Theta^T x$ is greater than $\Theta^T x$.

$$h(z) = \text{cost}_1 = \max(0, k(1-z)) = \begin{cases} 0 \text{ if } z \geq 1 \\ \\ 1 - z \text{ if } z < 1 \end{cases} \tag{2.28}$$

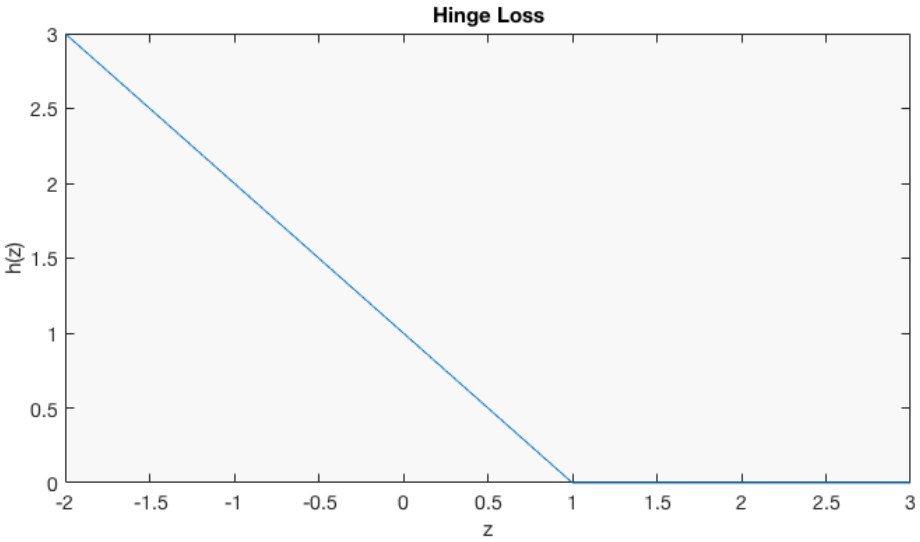


Figure 2.7: The Hinge Loss function for the first term, cost_1, in the cost function in equation 2.31.

$$h(z) = \text{cost}_0 = \max(0, k(1+z)) = \begin{cases} 0 \text{ if } z < -1 \\ 1 + z \text{ if } z \geq -1 \end{cases} \tag{2.29}$$

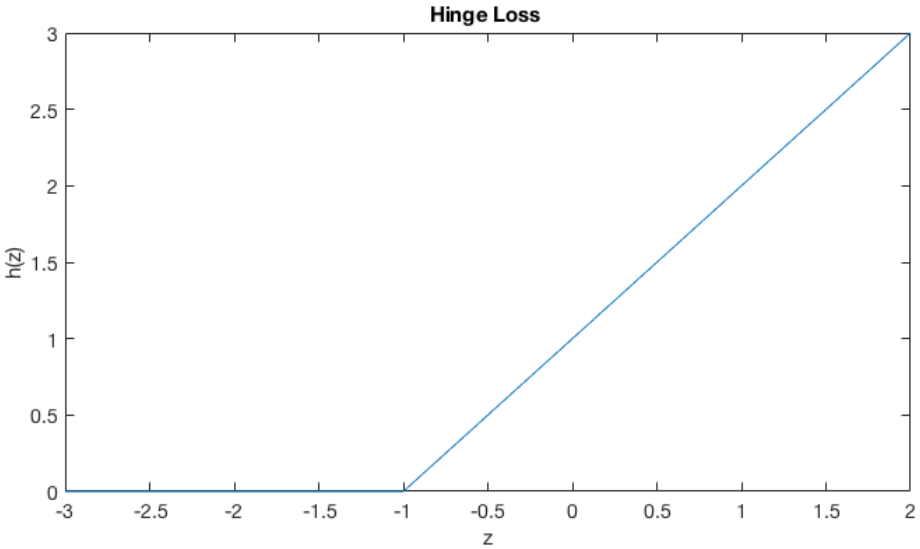


Figure 2.8: The Hinge Loss function for the second term, cost_0, in the cost function in equation 2.31.

Here k is a constant and cost_1 is the cost when classifying y = 1 and cost_0 is the cost when classifying y = 0. When replacing the terms for when y = 0 and y = 1 in equation 2.24 the cost function for SVM becomes

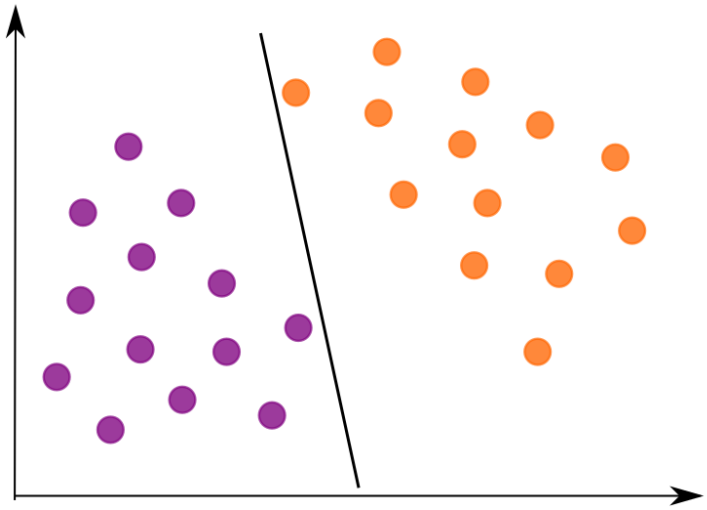


Figure 2.9: A linear SVM classifier.

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} y^{(i)} \text{cost}_1(\theta^T x^{(i)}) + (1 - y^{(i)}) \text{cost}_0(\theta^T x^{(i)}) + \frac{\lambda}{2m}\sum_{j=1}^{n} \Theta_j^2 \tag{2.30}$$

When using SVM it is usual to use C instead of λ as the regularisation constant. In addition we multiply by m, so the new cost function becomes

$$J(\theta) = C\sum_{i=1}^{m} y^{(i)} \text{cost}_1(\theta^T x^{(i)}) + (1 - y^{(i)}) \text{cost}_0(\theta^T x^{(i)}) + \frac{1}{2}\sum_{j=1}^{n} \Theta_j^2 \tag{2.31}$$

C is then equal to $\frac{1}{\lambda}$. This function will not give a probability, as function 2.24 did. This will either output 0 or 1, depending on the value of $h_\Theta(x)$.

$$h_\theta(x) = \begin{cases} 1 \text{ if } \Theta^T x \geq 0 \\ 0 \text{ if } \Theta^T x < 0 \end{cases}. \tag{2.32}$$

This gives a rather short margin on either side of the line, like in figure 2.9. Here the line touches very close to some of the points, and it's easy to see that this is not the optimal line to use when classifying between the two groups. Therefore we implement a larger margin.

$$h_\theta(x) = \begin{cases} 1 \text{ if } \Theta^T x \geq 1 \\ 0 \text{ if } \Theta^T x \leq -1 \end{cases}. \tag{2.33}$$

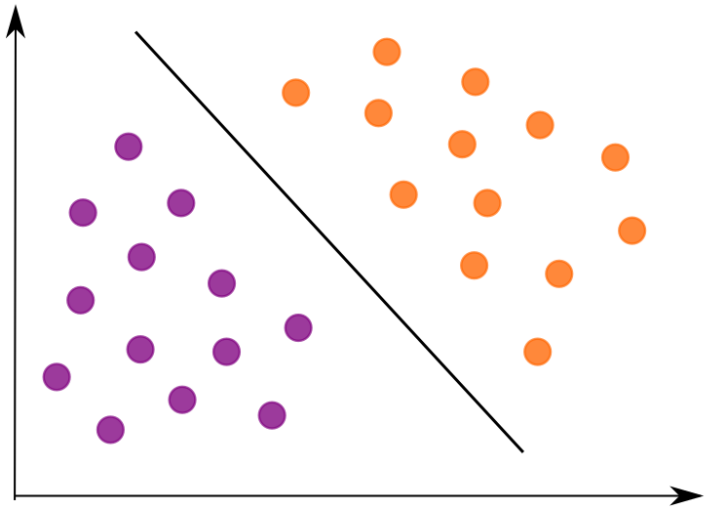


Figure 2.10: A large margin SVM classifier.

In figure 2.10 a large margin classifier is implemented. In this figure we can see that the line which draws the line between the two groups have a maximum distance from each group. To get a large margin, the regularisation constant C has to be large(inf).

Until now we have only looked at linear problems, problems were the solution is a straight line. This might not always be the case, and will rarely be the case in real problems. Therefore it is normal to implement kernels.

Lets say we have a point x, then we have some landmarks, $l^{(1)}, l^{(2)}, l^{(3)}$. Now we can calculate how close x are to the different landmarks. This gives us a new feature, f_1. The calculation of this feature is called a similarity function. There are several types of similarity functions, the Gaussian or RGB similarity function is the following

$$f_1 = \text{similarity}(x, l^{(i)}) = exp(-\frac{||x - l^{(i)}||}{2\sigma^2}). \tag{2.34}$$

This function has several useful properties. If x is very far from the landmark, then f_1 is zero, and the closer x is to the landmark, the closer f_1 is to one.

One way to chose the landmarks is to put them in same spot as the training exampels. Then you have m landmarks, as there are m training exampels. To get the parameters for θ the function to minimize is now the cost function, 2.31, but with $x^{(i)}$ substituted by $f^{(i)}$

$$\min_{\Theta} C \sum_{i=1}^{m} y^{(i)} \text{cost}_1(\Theta^T f^{(i)}) + (1 - y^{(i)}) \text{cost}_0(\theta^T f^{(i)}) + \frac{1}{2} \sum_{j=1}^{n} \Theta_j^2. \tag{2.35}$$

When using SVM on problems where there are several classes one method which can be used is the one vs all. When using one vs all you use the algorithm where y $\in$ 1,2,3,...,n with $\Theta^{(1)}$,$\Theta^{(2)}$,$\Theta^{(3)}$,...,$\Theta(n)$. You then calculate $(\Theta^{(i)})^T x$ for all the possibilities, the class, i, is then decided by which values yields the highest value for $(\Theta^{(i)})^T x$.

Chapter 3

Methods and Materials

Some of the sections are based on the equivalent sections in the spesialisation project[3] done during the fall of 2016.

3.1 Patient Cohort

The patient cohort used for the motion correction is the same as used by Teruel et al. in [34], but an additional 12 patients was included. This made a total retrospective cohort of 73 patients with locally advanced breast cancer. One patient had to be excluded based on the lack of DCE-MRI images, this made a subset of 72 patients included in the study. Patients who received neoadjuvant chemotherapy was not considered for this study. Patients with malingnant tumours were referred to surgery, and the tumour was hisotpathologically analysed. The criteria for defining the different types of breast cancer can be found in the article by Teruel [34].

For the coregistration a subsection of 18 patients was chosen, 6 patients with benign tumours, 6 patients with HER2- and 6 patients with HER2+. The information was provided by MR group at the Department of Physics at NTNU. The machine learning model made was based on this subsection.

3.2 MRI Protocol

The patients went through MRI imaging before neoadjuvant chemotherapy was started. The MRI scanner used was a 3-T scanner (Skyra; Siemens, Erlangen, Germany). The coil used was a

16-channel bilateral breast coil. A T2 weighted non-fat suppressed image was acquired early in the acquisition time. This sequence is a turbo spin echo sequence with TR/TE = 5500 / 118 ms. Section thickness 2.5 mm, flip angle 120°, acquisition matrix 256 x 256 and in plane resolution, 0.7 mm^2. T1-weighted DCE-MRI was acquired using a three-dimensional radiofrequency-spoiled gradient-echo sequence. The flip angle used was 15°; TR/TE = 5.82/2.18 ms; in plane resolution, 0.7 mm^2; acquisition matrix, 256 x 256; section thickness, 2.5 mm and 60 sections.

After the acquisition of a reference image, a bolus injection of 0.1 $\mathrm{mmol\,kg^{-1}}$ body weight of contrast agent was given. The contrast agent was gadolinium based (Dotarem; Guerbert, Bloomingtin, Ind) and was given at a rate of 2 $\mathrm{ml\,s^{-1}}$ followed by a 20 ml saline flush at the same injection rate. Then seven images were acquired with a temporal resolution of 1 min after the contrast injection.

3.3 Motion Correction and Coregistration

The software used to Motion correct and coregister images was CMTK, Computational Morphometry Toolkit [26]. CMTK is a MATLAB package which provides tools for motion correction and coregistration. In CMTK it is possible to specify the similarity function used, and the number of degrees of freedom (DoF). Both equation 2.13 equation 2.18 provides possible similarity functions which can be used by CMTK.

When motion correcting the images, both NCC and NMI was used, and three DoF, six DoF and nine DoF was tried with the NCC. With NMI only three DoF was tried.

Before the images could be analysed, they were coregistered to a common reference image. The reference image chosen was the T2 image without fat suppression. Images from 18 patients was co-registered. The group included 6 benign patients, 6 HER2- patients and 6 HER2+ patients. When the images were coregistered 6 DoF was used. In addition NCC was used as the similarity function.

3.4 ROIs

ROIs were made for three regions in the coregistered images. One ROI was made from air, one from the tumour and one which contained breast tissue. Breast tissue is here both fibroglandular tissue and fat tissue. There were also made ROIs which only contained fibroglandular tissue,

and ROIs which contained only fat. The ROIs were taken from the slice of the breast with most visible tumour on the subtraction image. A subtraction image is an image in the DCE series where the first, pre-contrast image is subtracted from the post-contrast images. This makes the tumour light up, and makes making the ROIs easy. For the test set a different slice, but still with visible tumour, was chosen. In some cases, where the tumour was large, some part of the tumour was chosen for the training set, and a different part for the test set.

3.5 Machine Learning

It was only ROIs from the co-registered DCE images which was added to the algorithm. The algorithm used was from LIBSVM, which is a free library which can be implemented in Matlab, python, C++, etc. In this project the Matlab implementation was used. LIBSVM has many useful functions and algorithms for support vector machine classification [35]. FITCSVM is a function which returns a SVM model. In this function the input is a table with data, and an array,Y, with the response. The table with data was made using intensity information from the pixels in the ROIs. In this case the table had 8 columns, one for every DCE image in the acquisition series, and the number of rows was given by how many pixels there was in any given training set. The array Y had the same number of rows as the table, and stated whether the pixel contained air, breast tissue, benign tumour or malignant tumour. The array Y therefore gave the class labels.

Since there were several classes, an one vs all approach was used. In one vs all a for loop is implemented which checks whether a given class is the same as the class for the pixel checked. This is done by calculating a value, and then check it for the different classes. Since the algorithm only takes two classes at a time, you have to implement a parameter called Class Names and set it to be either true or false. If the pixel checked then is the same class, you will get out true, otherwise false.

In addition the feature Standardise of the FITCSVM function was set to true. This leads to the table being standardised by dividing the columns by their standard deviation. The function used as the kernel was the Gaussian function 2.34.

The model was trained using a training set and tested using a test set. The training set and test set contained ROIs from different slices of the same patient. In addition we tried to included

patients which had not been used to train the model in the test set. The model was also tested using the training set as the test set.

Chapter 4

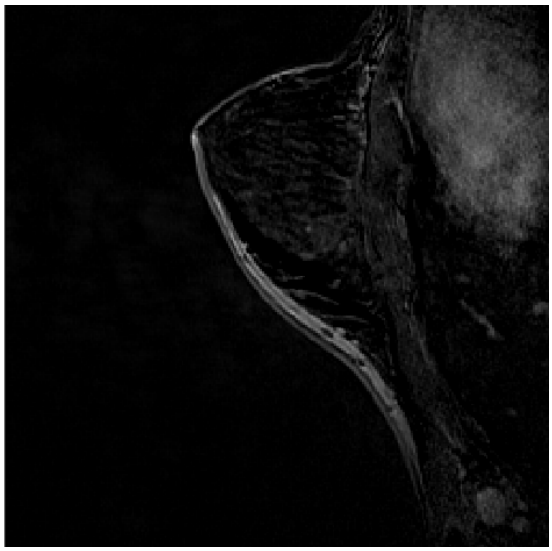
Results

4.1 Motion Correction

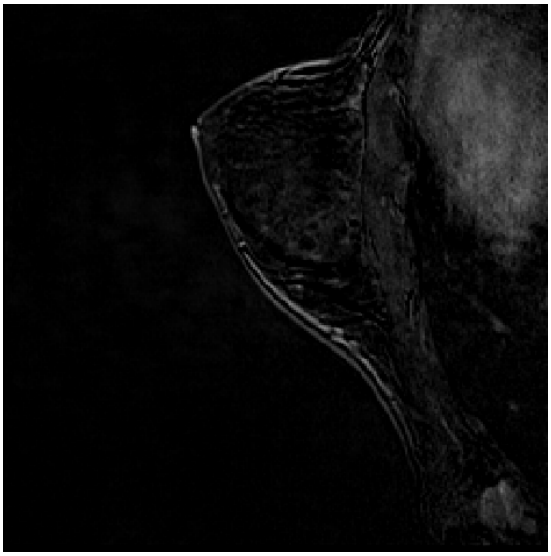
As a result from the experience of the specialisation project[3] during the fall 2016 we decided to try to increasing the number of DoF with NCC and to try a different motion correction algorithm for some of the patients where NCC yielded unsatisfactory results. The results from the specialisation project showed that not all motion was corrected. This is shown in the subtraction images in figure 4.1, where the pre-contrast DCE image has been subtracted from the post-contrast DCE images. The motion is specially clear around the boarders of the breast in the original image, see figure 4.1a, where thick white lines are visible. In the motion corrected breast where NCC with three DoF, see figure 4.1b, is used there are less visible lines, but they are still apparent.

In figure 4.3 the dynamic subtraction images where the original images are subtracted from motion corrected images using NCC and three DoF are shown for patient 26. In these images it is possible to see how the motion varies during the acquisition series. Since there are some clear lines around the edges of the breast, it means that some of the motion from the original images has been removed, and there is less motion in the new set of images. This was the standard method we used in the specialisation project. As we see from figure 4.1 this did not remove all the motion, and there was room for improvement.

The first change we applied to the motion correction algorithm in order to yield more satisfactory results was to increase the number of DoF. In the standard method we used NCC with 3 DoF, now we increased it to six DoF. The result can be seen in figure 4.4. The images looks



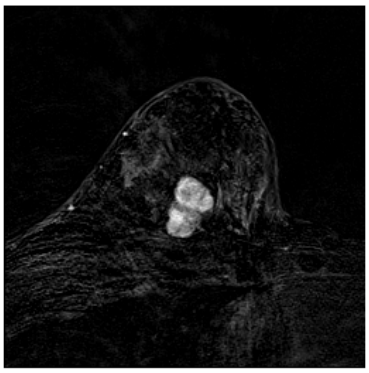
(a) Patient 99 original image, section 25 frame 5



(b) Patient 99 motion corrected image, section 25 frame 5

Figure 4.1: Patient 99 original image and motion corrected image. The images is from section 25, the forth image frame after contrast injection. There is less visible lines around the border of the breast after the motion correction, but the lines are still visible.

similar to the images in figure 4.1, but it seems as a bit more motion is removed when increasing the DoF to six. This difference can be seen more easily in figure 4.2 where the edges are a bit less visible in figure 4.2b than in figure 4.2a.

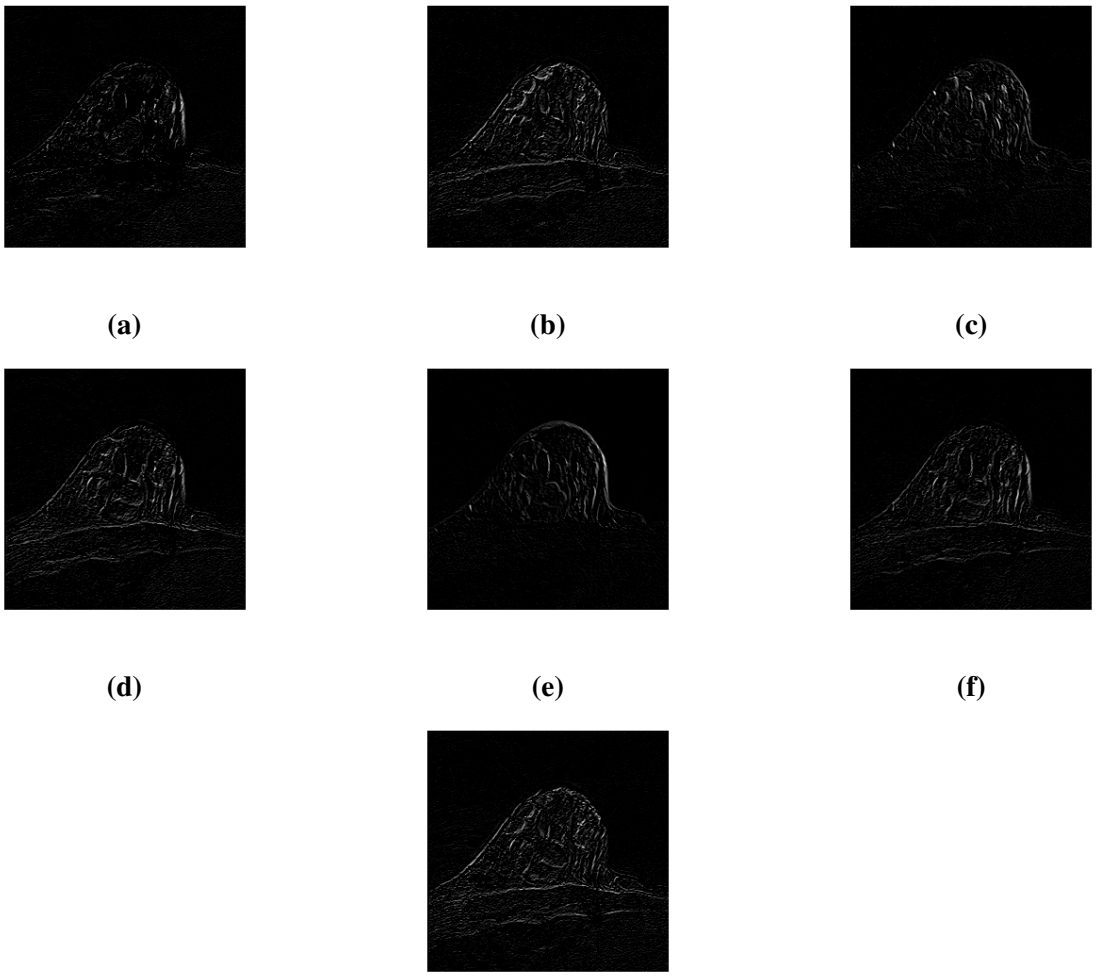


(a) Motion corrected using NCC and 3 DoF.

(b) Motion corrected using NCC and 6 DoF.

Figure 4.2: Motion corrected images of patient 26, slice 33, second post-contrast image.

The next step was to increase the number of DoF further, to nine. The result of this can be seen in figure 4.5. When looking at the difference between six and nine DoF, it is not visible. When looking at the running time as well for the algorithm when using 9 DoF, the difference



(a) **(b)** **(c)**

(d) **(e)** **(f)**

(g)

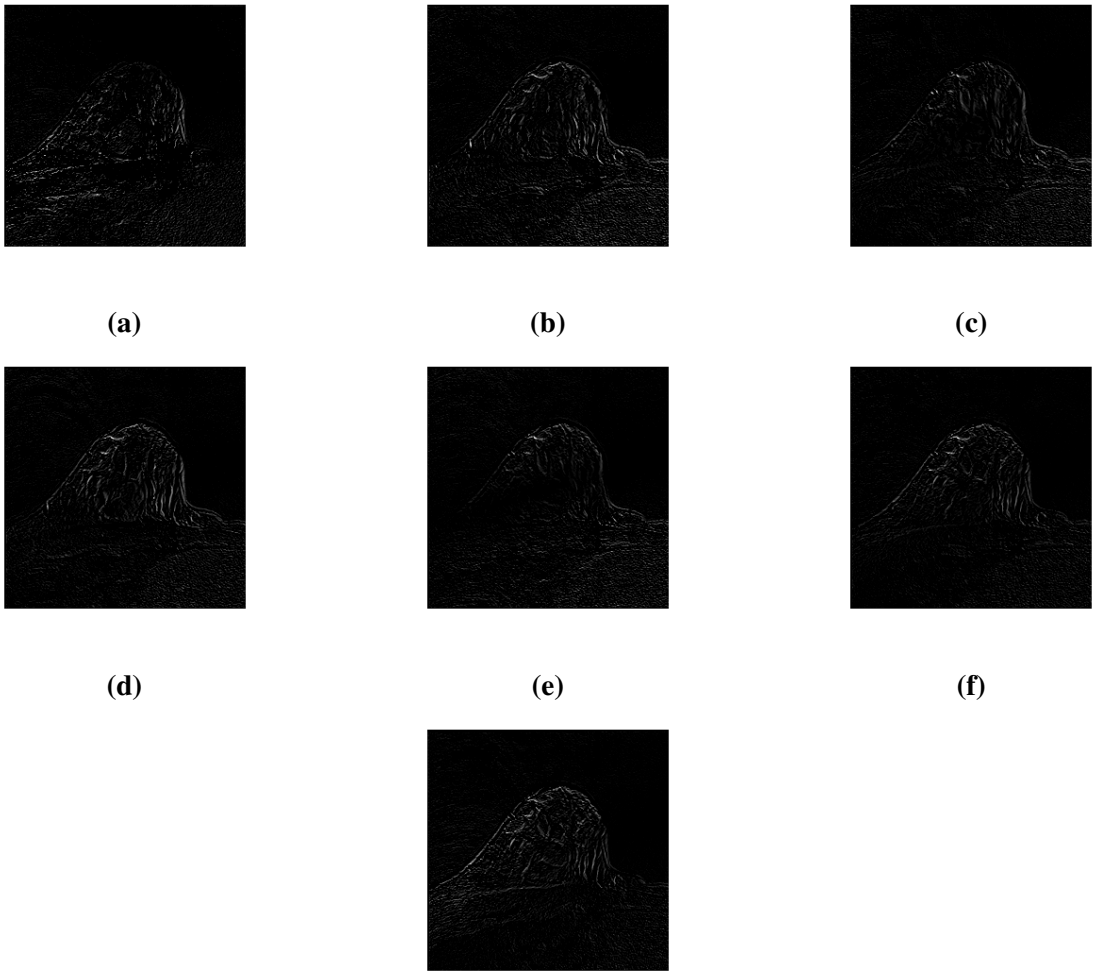
Figure 4.3: Subtraction images of original images and images motion corrected with NCC and 3DoF from patient 26, slice 33. Image a-g are post-contrast images, where a is the first. The images are taken with a 1 minute interval.

might be to small to be worth the increased computational running time.

Finally another algorithm was tried, NMI. When implementing the algorithm 3 DoF was used. The result of this for patient 26 can be seen in figure 4.6. There is about the same amount of corrected tissue as for NCC with 3 DoF.

4.2 Coregistration

The co-registration was done with a T2 weighted non-fat suppressed image as a reference. For patient 27 the reference T2 image can be seen i figure 4.7a. This image is usually used for

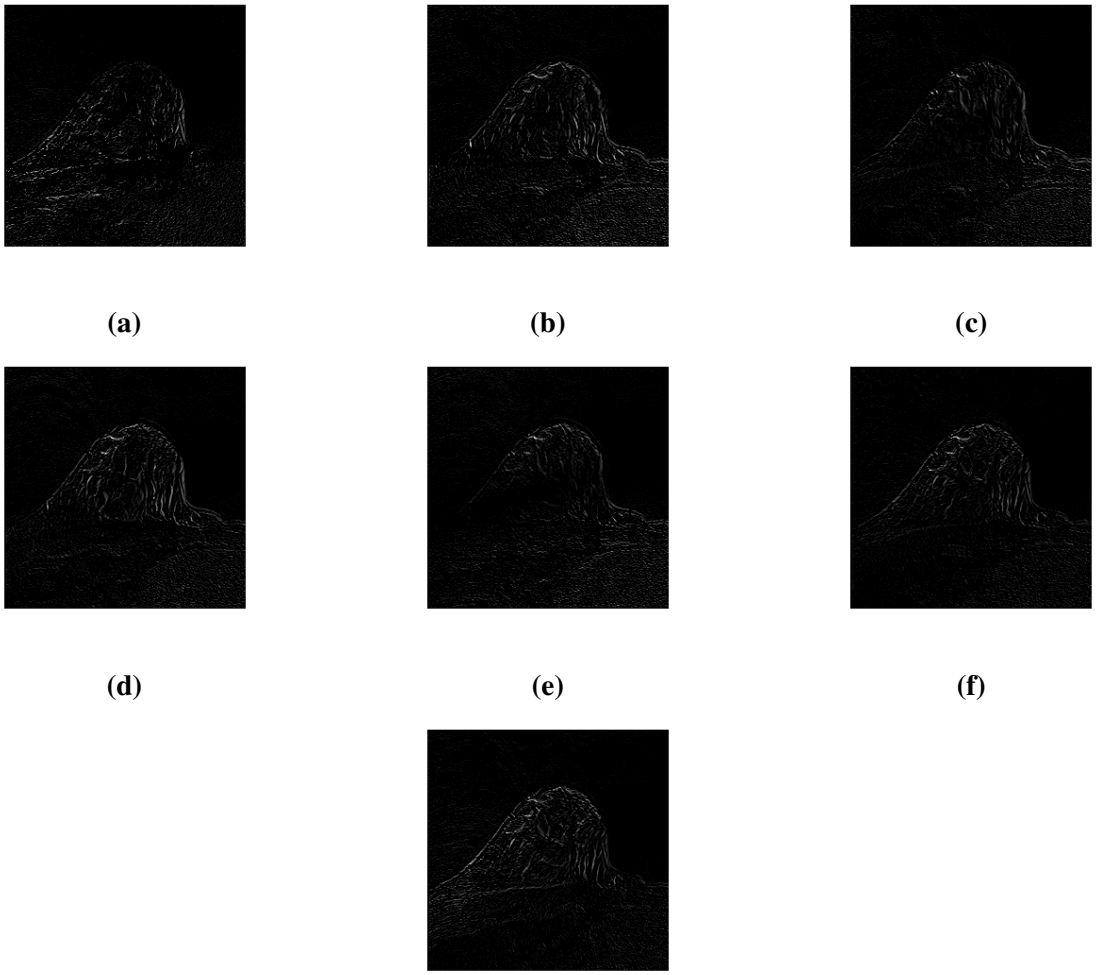


(a) **(b)** **(c)**

(d) **(e)** **(f)**

(g)

Figure 4.4: Subtraction images of original images and images motion corrected with NCC and 6 DoF from patient 26, slice 33. Image a-g are post-contrast subtraction images, where a is the first image taken, and g the last. The images are taken with a 1 minute interval.

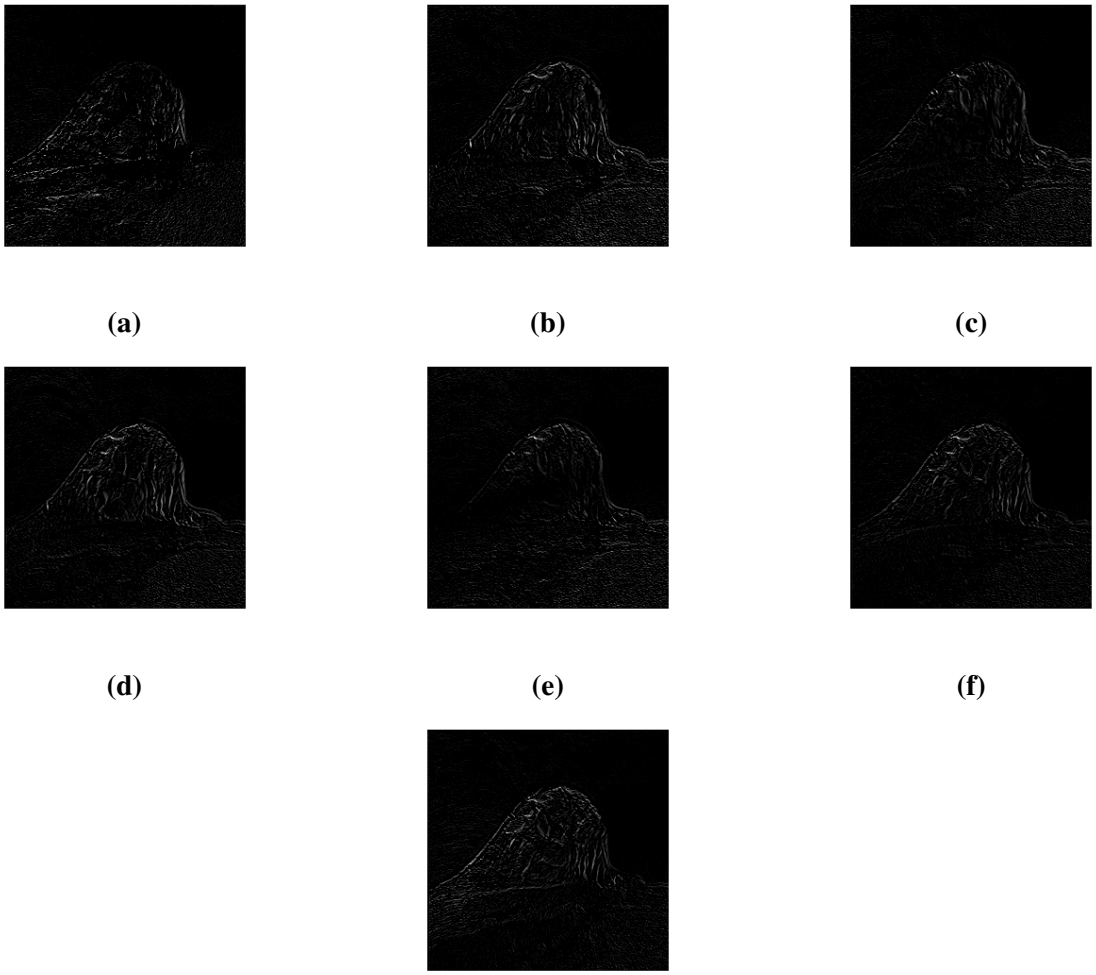


(a) **(b)** **(c)**

(d) **(e)** **(f)**

(g)

Figure 4.5: Subtraction images of original images and images motion corrected with NCC and 9 DoF from patient 26, slice 33. Image a-g are post-contrast images, where a is the first image and g the last. The images are taken with a 1 minute interval.



(a) **(b)** **(c)**

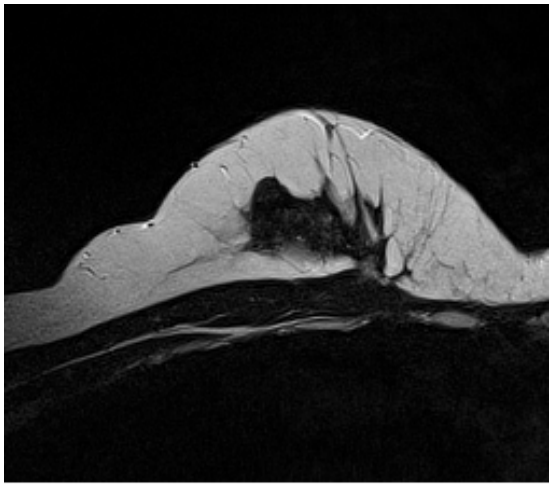
(d) **(e)** **(f)**

(g)

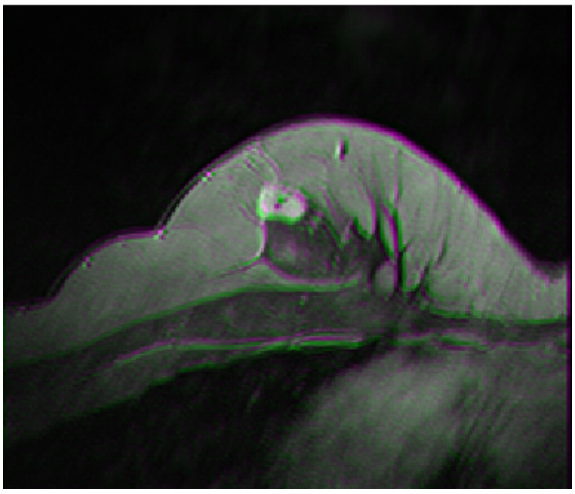
Figure 4.6: Subtraction images of original images and images motion corrected with NMI and 3 DoF from patient 26, slice 33. Image a-g are post-contrast images, where a is the first image and g the last. The images are taken with a 1 minute interval.

anatomical reasons, and as we can see, the tumour is not clearly visible. For patient 27 slice 33 was chosen as it was the slice with the most visible tumour. In figure 4.7b the motion corrected DCE image and the coregistered motion corrected DCE image are layered over each other to show the difference. The coregistered motion corrected image is green and the motion corrected image is magenta. Where the colours are visible there has been movement between the T2 reference image and the motion corrected image, so the coregistered image has been shifted to adjust for the movement. There colours are specially visible along the edges of the breast.

The motion corrected images used in the coregistration was motion corrected using NCC and 3 DoF. As we can see from figure 4.7b it is possible to discern the position of the tumour from these images, but it is not possible to decide whether the tumour are benign or malignant. The coregistration was done to make it possible to compare images from different MRI modalities, like diffusion images.



(a) T2 reference image.



(b) Layered image.

Figure 4.7: Patient 27, slice 33. The layered image is of the motion correction image and the coregistered motion corrected image from the second post-contrast image. The images are coloured magenta and green respectively.

4.3 ROIs

ROIs was made from all the included patients. To find the ROI for the tumour the subtraction image, where the pre-contrast image is subtracted from an image later in the imaging series, was used. Figure 4.8 was used for patient 27. This is the subtraction image for slice 33, and as

we can see the tumour lights up in the image. In figure 4.9a the ROI for the tumour which was used for patient 27 is shown. The ROI can be seen as a light grey square inside the tumour. In figure 4.9b the ROI for air is shown for patient 27. The ROI is the light grey square. In figure 4.9c the ROI for breast tissue for patient 27 is shown. In image 4.10a, 4.10b and 4.10c similar images is shown for patient 30. In this patient slice number 37 was used, as it was the slice which contained the most tumour.

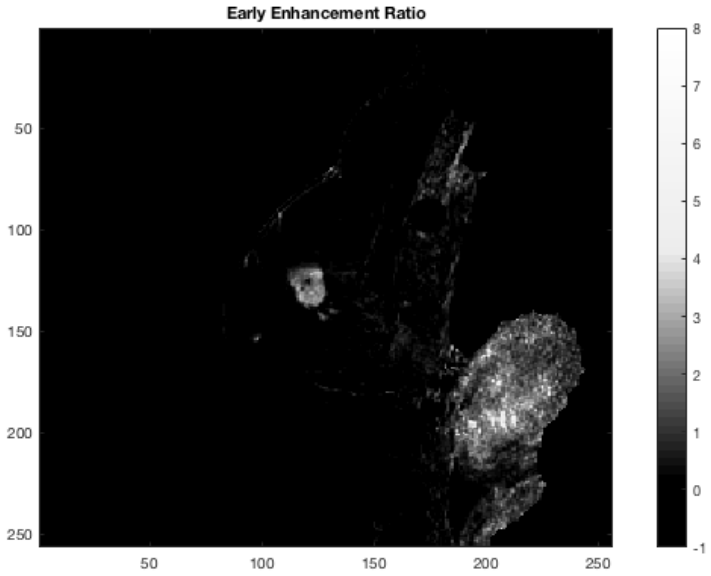


Figure 4.8: The second post contrast injection image, where the pre-contrast image has been subtracted. The tumour is easily visible in the middle of the figure. The part lighting up the lower right part of the figure might be the heart.

4.4 Machine Learning

The first case tried was with one patient with a benign tumour. Patient 26 was chosen, and the slice chosen to make the training set was number 32. Slice 33 was used to make the test set. ROIs was made for air, tumour and breast tissue. The model made from the training set was then tried on the test set. The percentage of correctly classified pixels in the table was found to be 99.7 %.

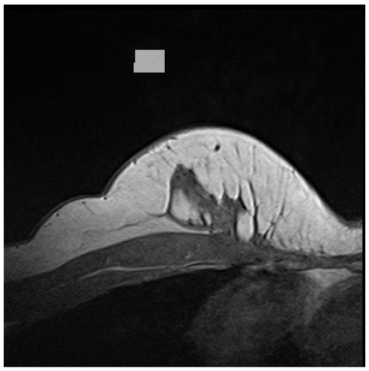
At first we tried to make the model using ROIs for fibroglandular tissue, fat tissue, air and tumour tissue. Since it was hard to make the ROIs precise, specially for the fibroglandular tissue, this confused the model. The result of this was a poor classification rate, of about 70%,

compared to when fibroglandular tissue and fat tissue was put in the same ROIs, resulting in a classification rate of 99.7%.

Then a patient with a malignant tumour was added to the training set. The training set now contained ROIs from both a benign tumour and a malignant tumour. The patient chosen was number 27, which has a HER2- tumour. When ROIs from another slice from this patient was added to the test set, the model correctly classified 97.36% of the pixels. When testing the model using the training set as the test set, the model correctly classified 97.58% of the pixels.

The next step was to add a HER2+ patient to the training and test set. Until now the classification had only been between benign and malignant, but now we tried to see whether the model would be able to classify the sub type of malignant tumour correctly as well. When adding an additional patient to the test set, which has not been used to train the model, the classification rate dropped to 90% of the pixels in the test set. When investigating which pixels the model had trouble classifying, it turned out that the model could not discern between the HER2+ and HER2- pixels. This cross test proved that more features are needed to train the model to make it able to discern between HER2- and HER2+ tumours.

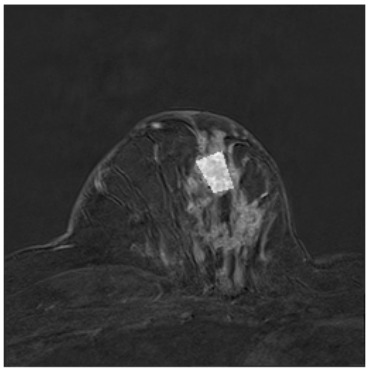
(a) The ROI for the tumour in patient 27. The ROI is the light grey square in the middle of the tumour.



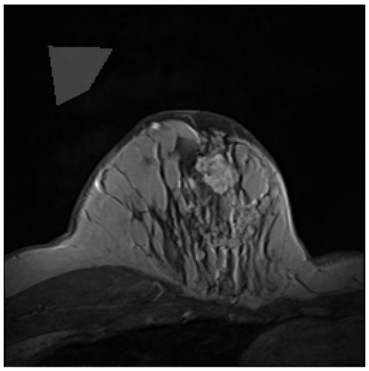
(b) The ROI for air in patient 27. The ROI is the light grey square in the black part of the figure.

(c) The ROI for the breast tissue in patient 27. The ROI is the light grey square in the middle of the breast.

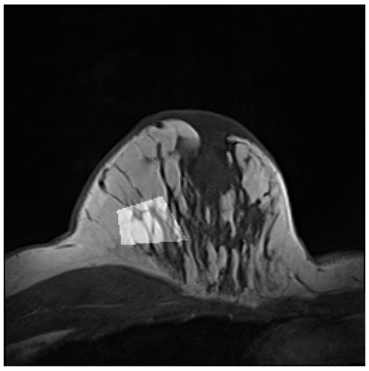
Figure 4.9: ROIs for patient 27. The difference in contrast results from the images being from different parts of the DCE acquisition series.



(a) The ROI for the tumour in patient 30. The ROI is the light grey square in the middle of the tumour.



(b) The ROI for air in patient 30. The ROI is the light grey square in the black part of the figure.



(c) The ROI for the breast tissue in patient 30. The ROI is the light grey square in the middle of the breast.

Figure 4.10: ROIs for patient 30. The difference in contrast results from the images being from different parts of the DCE acquisition series.

Chapter 5

Discussion

5.1 Motion Correction

From the specialisation project [3] in the fall some questions was not answered. One of them was whether an increase in the number of DoF, or using another similarity function, might improve the motion correction in cases where NCC and 3 DoF yielded unsatisfactory results. In this project we chose to evaluate whether we could improve the motion correction further. First the number of DoF was increased from three to six for patients which had visible motion even after the initial motion correction, see figure 4.1b and 4.3. The similarity function used was still NCC. This yielded improved results, which can be seen in figure 4.2 and figure 4.4. Then the number of DoF were further increased to nine. A we can see from figure 4.5 this figure looks similar to figure 4.4. The reason for the rather small change between six and nine DoF might be that there are only small amounts or no need for non-uniform scaling to correct for the motion in the breast during the acquisition time. An additional point against using nine DoF is the increased amount of computational time required to motion correct using nine DoF and NCC.

An evaluation was done to not further increase the number of DoF, as the computational time would continue to increase and it is likely that the images would not improve by a large degree, as the breast is not a rigid body. When using DoF to correct the images we account for motion of rigid bodies. An example of a rigid body is the head, the head cannot be easily deformed. The breast on the other hand can, which makes the transformations not only global. In the breast there will be some deformation, which indicates that there are likely local transformations as

well. So if more motion correction is to be applied to these images it might be an idea to consider correcting for these local transformation, rather than more global transformations.

The last change applied to the motion correction was to try another similarity function. The selected function was NMI, see equation 2.18, and this yielded a slightly different image than with NCC, see figure 4.6. The difference between images corrected with NMI and with NCC seems to be small. There are some differences in which part of the breast receives the most correction, this is visible when comparing figure 4.3a and 4.6a. When looking at these to figures it seems that NMI corrects less of the motion at the right side of the breast and NCC corrects less of the motion at the left side of the breast.

Since the difference between using NMI and NCC are minor, and more of the motion is corrected using NCC with six DoF than NMI with 3 it was decided that the patients with unsatisfactory standard motion correction would be motion corrected using six DoF and NCC.

5.2 Coregistration

The next part of the project was the coregistration of DCE images with a reference image. In this project we choose to use a T2 weighted image as our reference. The reason for this choice was that we had T2 weighted images without fat-suppression for all patients. T2 weighted images is taken for all patients as one of the first acquisition series. This is often done for anatomical reason, and is therefore a good choice for the reference image. In the beginning we considered using the T2 fat suppressed image as the reference image, but we decided against it. The main reason for this was because we did not have T2 fat suppressed images for all the patients. There maybe many reasons for this, one being the patient being claustrophobic and decreasing the acquisition time might be needed. The series with T2 fat suppressed images is one of the last series to be done in the acquisition series and is therefore it is not always done if the acquisition time is to be decreased. Another reason why this is not always done might be because the patient moved to much during the acquisition to yield good results, and therefore the acquisition was ended early.

The coregistration was done for 18 patients. When choosing the patients we chose patients with as little motion as possible. This was done to minimise the effect of motion in this pilot study. In patients with a lot of motion the coregistration, and then the machine learning, would

probably not work as well as it did. In the future they would have to be included as well, so then a better motion correction algorithm might have to be found. The advantage of not including them in this early project was that the error from motion of the patient was reduced. Since the patients nearly did not move, there was not much movement between the different acquisition series, and therefore the difference between the DCE images and the T2 weighted images were small. A small difference means that we can assume that every pixel is registered correctly, and that the development in intensity is correct. The movement of a patient could lead to misclassification of the pixels, specially the pixels around the border of the tumour. The reason for this is that a pixel might be inside the tumour in the first images, and then be registered as a pixel outside the tumour in the later images.

5.3 ROIs

The ROIs made from the DCE MRI images were from air, breast tissue and tumour tissue. We first tried to make ROIs from fat tissue and from fibroglandular tissue. But since I am not a physician, it was hard to see the difference. In addition there are sometimes muscle tissue in the images(lying outside the ribs). For this reason we decided that all breast tissue (fat and fibroglandular tissue) should be categorised as one type of tissue, breast tissue. This made making the ROIs much easier, and it improved the SVM model. The first attempts of making the model used ROIs from air, tumour tissue, fibroglandular tissue and fat tissue. But with these classes the model was only able to classify about 70% of the pixels correctly. The reason for this was probably that fat tissue had been included in the fibroglandular ROIs as it was hard to only include fibroglandular tissue in the ROI. When combining the fibroglandular tissue and fat tissue in one class the classification rate increased to 99.7% when only benign tumours where included.

The tumour was easily found using subtraction images from the DCE acquisition series. In these images the tumour lights up, and makes it is easy to make a ROI. However it is not possible to decide whether the tumour is benign or malignant from the subtraction images.

5.4 Machine Learning

There might be a disadvantage of classifying all breast tissue as one type of tissue and not several as it actually is. One disadvantage is that the machine learning model developed might be confused. Since fat and fibroglandular tissue has different properties. Fibroglandular tissue is more dense than fat. Breast fat has a density of approximately 911 $\mathrm{kg\,m^{-3}}$ and the breast gland has a density of 1041 $\mathrm{kg\,m^{-3}}$ [36]. The T2 relaxation time for fat is 52.96 ms and 54.36 ms for glandular tissue [37]. This values implies that the contrast uptake and intensities of the fat tissue and the fibroglandular / breast gland tissue might have quite different properties. The density of fibroglandular tissue is actually 14 % higher than the density of fat. This might lead the machine learning model to classify pixels which is actually tumour pixels as breast tissue pixels, specially if the tumour is benign, so the contrast uptake is not as high as in malignant tumours. The advantage of classifying them together was seen in the previous section, the ROIs has to be made very carefully if the model is not to be confused.

When testing the model using a test set the correctly classified rate was 97.358% when including both benign and malignant tumours. It is interesting to also look at the rate when using the training set as the test set. If the model achieves a rate of 100 % it shows that there is a generalisation problem, and that the model overfits the data in the training set. Since the value we achieved when using the training set as the test set was 97.58%, it is likely that our model does not overfit the data.

When including classification between not only benign and malignant tumours but also including subtypes of malignant tumours the classification rate dropped to about 90% on a cross test, where a patient which had not been included in the training set was included in the test set. The model was then trained using both benign tumours, HER2+ tumours and HER2- tumours. When looking at the wrongly classified pixels it turned out that most of them were inside the tumour area, and that a classification rate of 90% in this case is unsatisfactory since the model clearly could not discern between HER2+ and HER2- tumours.

To increase the classification rate additional features has to be included in the matrix on which the model is based. Examples of features which could be included is the signal enhancement ratio, wash out and the values from ROIs from a T2 weighted image. Wash out and enhancement ratio are parameters which are calculated from parameters already existing in the matrix, so it is not certain that they will have any influence. The T2 weighted images on the

other hand provides new information, and might increase the classification rate. Other features which could improve the classification rate would be to include data from other MRI modalities, like diffusion weighted images (DWI). From DWI it is possible to find other parameters than from DCE, and these parameters might make it easier to classify between different kinds of malignant tumours.

Chapter 6

Conclusion

Motion correction was redone for some of the patients. The increase of number of DoF turned out to improve the motion correction for patients where 3 DoF was not enough. Using another similarity function did not improve the motion correction as much. Then the 18 chosen images were coregistered to a non-fat suppressed T2 weighted MRI image. From some of the coregistered images ROIs was made for air, breast tissue and tumour tissue. The intensity information from every pixel in the ROIs was then put into a table which was then used as the input in a SVM classifier. The model made by the SVM classifier was then tried on a test set. The result when testing only on benign tumours was 99.7% and when including malignant tumour it decreased to 97.4%.

6.1 Future Work

Future work include improving the SVM model to distinguish between different kinds of malignant tumours. This should be done by including information from other MRI modalities and by including training data from an increased number of patients.

Bibliography

[1] Christiane Katharina Kuhl, Peter Mielcareck, Sven Klaschik, Claudia Leutner, Eva Wardelmann, Jürgen Gieseke, and Hans H Schild. Dynamic breast mr imaging: Are signal intensity time course data useful for differential diagnosis of enhancing lesions? 1. *Radiology*, 211(1):101–110, 1999.

[2] Ahmedin Jemal, Freddie Bray, Melissa M Center, Jacques Ferlay, Elizabeth Ward, and David Forman. Global cancer statistics. *CA: a cancer journal for clinicians*, 61(2):69–90, 2011.

[3] Ingrid Quist-Løkken. Classification of breast cancer based on multiparametric mri. Unpublished specialisation project, 12 2016.

[4] Daniel Rueckert, Luke I Sonoda, Carmel Hayes, Derek LG Hill, Martin O Leach, and David J Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *IEEE transactions on medical imaging*, 18(8):712–721, 1999.

[5] Carla Boetes, Jelle O Barentsz, Roel D Mus, RF Van Der Sluis, LJ van Erning, JH Hendriks, Roland Holland, and SH Ruys. Mr characterization of suspicious breast lesions with a gadolinium-enhanced turboflash subtraction technique. *Radiology*, 193(3):777–781, 1994.

[6] Elisabeth V Heiberg, William H Perman, Virginia M Herrmann, and Christina G Janney. Dynamic sequential 3d gadolinium-enhanced mri of the whole breast. *Magnetic resonance imaging*, 14(4):337–348, 1996.

[7] WA Kaiser and E Zeitler. Mr imaging of the breast: fast imaging sequences with and without gd-dtpa. preliminary observations. *Radiology*, 170(3):681–686, 1989.

[8] AR Padhani. Dynamic contrast-enhanced mri studies in human tumours. *The British journal of radiology*, 72(857):427–431, 1999.

[9] Atle Bjørnerud. The physics of magnetic resonance imaging. Compendium, 3 2008.

[10] Fahmi Khalifa, Ahmed Soliman, Ayman El-Baz, Mohamed Abou El-Ghar, Tarek El-Diasty, Georgy Gimel'farb, Rosemary Ouseph, and Amy C Dwyer. Models and methods for analyzing dce-mri: a review. *Medical physics*, 41(12):124301, 2014.

[11] Paul S Tofts. T1-weighted dce imaging concepts: modelling, acquisition and analysis. *signal*, 500(450):400, 2010.

[12] Anwar R Padhani and Janet E Husband. Dynamic contrast-enhanced mri studies in oncology with an emphasis on quantification, validation and human studies. *Clinical radiology*, 56(8):607–620, 2001.

[13] Zishan A Haroon, Kevin G Peters, Charles S Greenberg, and Mark W Dewhirst. Angiogenesis and oxygen transport in solid tumors. In *Antiangiogenic agents in cancer therapy*, pages 3–21. Springer, 1999.

[14] Robert W Brown, Y-C Norman Cheng, E Mark Haacke, Michael R Thompson, and Ramesh Venkatesan. *Magnetic resonance imaging: physical principles and sequence design*. John Wiley & Sons, 2014.

[15] Fred W Flickinger, Jerry D Allison, Richard M Sherry, and John C Wright. Differentiation of benign from malignant breast masses by time-intensity evaluation of contrast enhanced mri. *Magnetic resonance imaging*, 11(5):617–620, 1993.

[16] Habib Rahbar and Savannah C Partridge. Multiparametric mr imaging of breast cancer. *Magnetic resonance imaging clinics of North America*, 24(1):223–238, 2016.

[17] Paul S Tofts. Modeling tracer kinetics in dynamic gd-dtpa mr imaging. *Journal of Magnetic Resonance Imaging*, 7(1):91–101, 1997.

[18] Paul S Tofts, Gunnar Brix, David L Buckley, Jeffrey L Evelhoch, Elizabeth Henderson, Michael V Knopp, Henrik BW Larsson, Ting-Yim Lee, Nina A Mayr, Geoffrey JM Parker, et al. Estimating kinetic parameters from dynamic contrast-enhanced t 1-weighted mri of a diffusable tracer: standardized quantities and symbols. *Journal of Magnetic Resonance Imaging*, 10(3):223–232, 1999.

[19] Matthias C Schabel, Glen R Morrell, Karen Y Oh, Cheryl A Walczak, R Brad Barlow, and Leigh A Neumayer. Pharmacokinetic mapping for lesion classification in dynamic breast mri. *Journal of Magnetic Resonance Imaging*, 31(6):1371–1378, 2010.

[20] Bernd Girod. *Signals and Systems*. John Wiley & Sons, 2001.

[21] Paul Viola and William M Wells. Alignment by maximization of mutual information. In *Computer Vision, 1995. Proceedings., Fifth International Conference on*, pages 16–23. IEEE, 1995.

[22] André Collignon, Frederik Maes, Dominique Delaere, Dirk Vandermeulen, Paul Suetens, and Guy Marchal. Automated multi-modality image registration based on information theory. In *Information processing in medical imaging*, volume 3, pages 263–274, 1995.

[23] Jay B West, J Michael Fitzpatrick, Matthew Y Wang, Benoit M Dawant, Calvin R Maurer Jr, Robert M Kessler, Robert J Maciunas, Christian Barillot, Didier Lemoine, Andre MF Collignon, et al. Comparison and evaluation of retrospective intermodality image registration techniques. In *Medical Imaging 1996*, pages 332–347. International Society for Optics and Photonics, 1996.

[24] Frederik Maes, Andre Collignon, Dirk Vandermeulen, Guy Marchal, and Paul Suetens. Multimodality image registration by maximization of mutual information. *IEEE transactions on medical imaging*, 16(2):187–198, 1997.

[25] Colin Studholme, Derek LG Hill, and David J Hawkes. An overlap invariant entropy measure of 3d medical image alignment. *Pattern recognition*, 32(1):71–86, 1999.

[26] Torsten Rohlfing. *User Guide to the Computational Morphometry Toolkit*. Neuroscience Program, SRI International, Menlo Park, CA, 3 2011. An optional note.

[27] Nils J Nilsson. *Introduction to machine learning. An early draft of a proposed textbook*. Citeseer, 1996.

[28] Jeffrey M Stanton. *Introduction to data science*. 2013.

[29] Shai Shalev-Shwartz and Shai Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

[30] David Barber. *Bayesian reasoning and machine learning*. Cambridge University Press, 2012.

[31] Arthur L Samuel. Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3(3):210–229, 1959.

[32] Vijay Kotu and Bala Deshpande. *Predictive analytics and data mining: concepts and practice with rapidminer*. Morgan Kaufmann, 2014.

[33] Jason DM Rennie and Nathan Srebro. Loss functions for preference levels: Regression with discrete ordered labels. In *Proceedings of the IJCAI multidisciplinary workshop on advances in preference handling*, pages 180–186. Kluwer Norwell, MA, 2005.

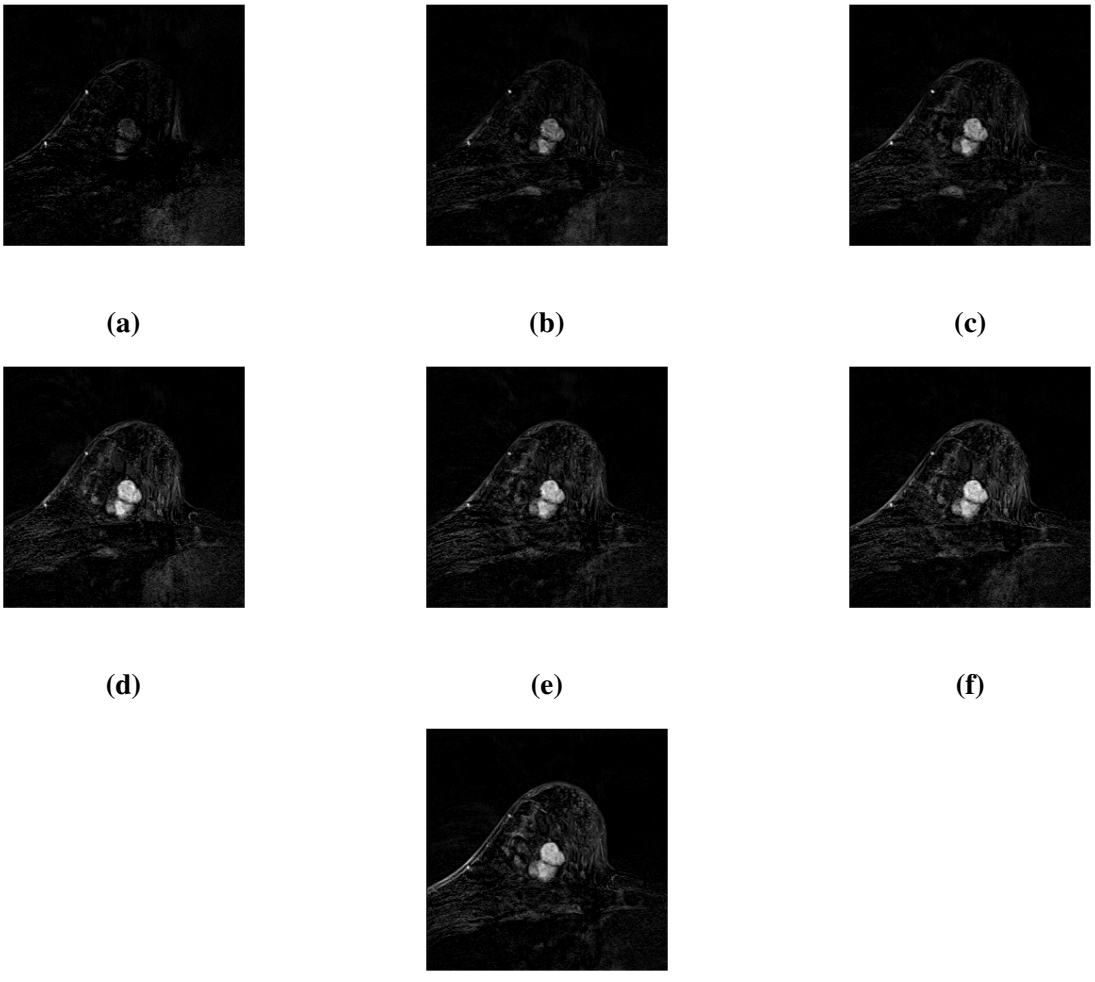
[34] Jose R Teruel, Pål E Goa, Torill E Sjøbakk, Agnes Østlie, Hans E Fjøsne, and Tone F Bathen. A simplified approach to measure the effect of the microvasculature in diffusion-weighted mr imaging applied to breast tumors: Preliminary results. *Radiology*, page 151630, 2016.

[35] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, et al. A practical guide to support vector classification. 2003.

[36] PA Hasgall, E Neufeld, MC Gosselin, A Klingenböck, and N Kuster. It'is database for thermal and electromagnetic parameters of biological tissues. *Version 3.0*, 2015.

[37] Rebecca Rakow-Penner, Bruce Daniel, Huanzhou Yu, Anne Sawyer-Glover, and Gary H Glover. Relaxation times of breast tissue at 1.5 t and 3t measured using ideal. *Journal of Magnetic Resonance Imaging*, 23(1):87–91, 2006.

Appendix

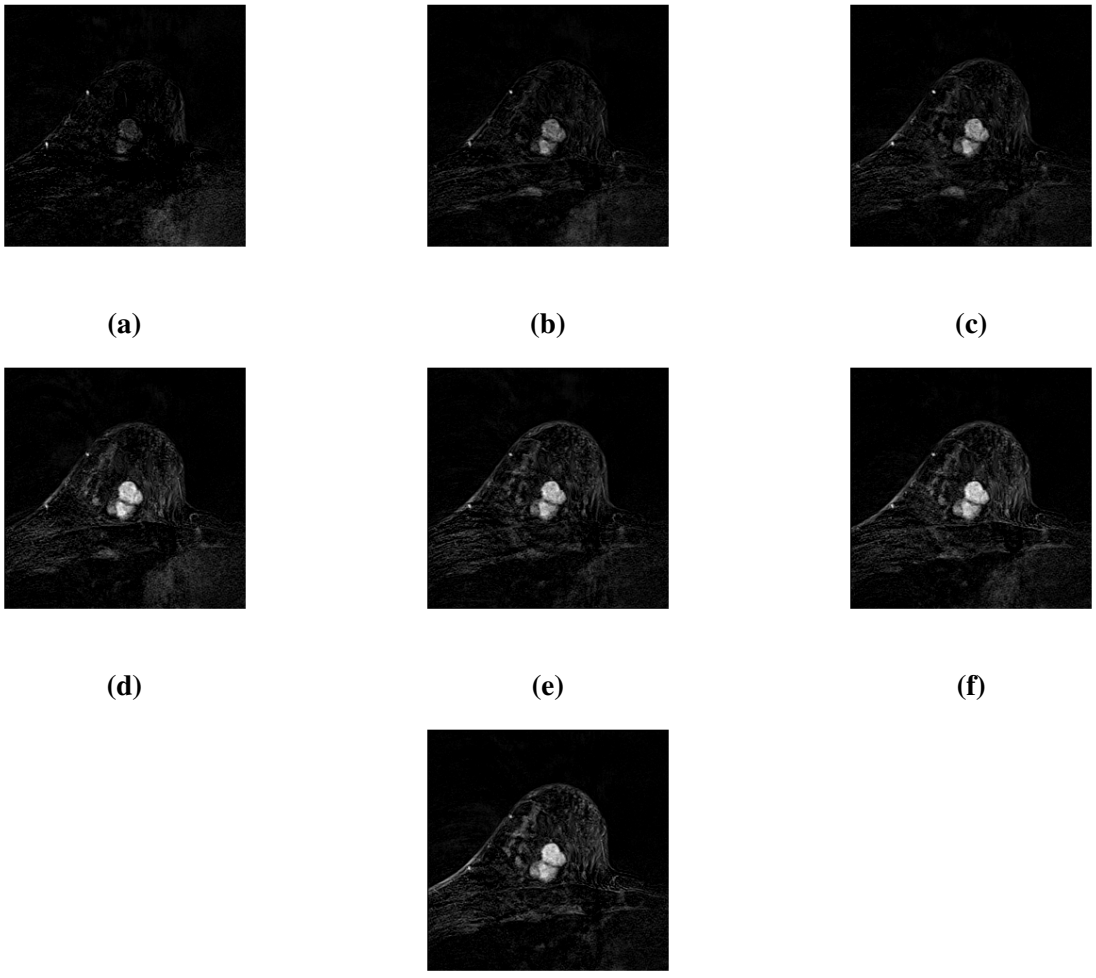


(a) **(b)** **(c)**

(d) **(e)** **(f)**

(g)

Figure 6.1: Original images for patient 26 without motion correction.



(a) **(b)** **(c)**

(d) **(e)** **(f)**

(g)

Figure 6.2: Images from patient 26 motion corrected with NCC and 3 DoF.