

**Automated Analysis of
Magnetic Resonance Imaging
of the Breast**

Albert Gubern-Mérida

This book was typeset by the author using L^AT_EX₂_ε.

Copyright © 2015 by Albert Gubern-Mérida. All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the author.

ISBN 978-94-6259-583-5

Printed by Ipskamp Drukkers, Nijmegen.

Automated Analysis of Magnetic Resonance Imaging of the Breast

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR AAN DE RADBOUD UNIVERSITEIT
NIJMEGEN OP GEZAG VAN DE RECTOR MAGNIFICUS PROF. DR. TH. L. M.
ENGELLEN, VOLGENS BESLUIT VAN HET COLLEGE VAN DECANEN EN TER
VERKRIJGING VAN DE GRAAD VAN DOCTOR AAN DE UNIVERSITAT DE GIRONA OP
GEZAG VAN DE RECTOR DR. S. BONET MARULL, IN HET OPENBAAR TE
VERDEDIGEN OP VRIJDAG 24 APRIL 2015 OM 12.30 UUR PRECIES

DOOR

Albert Gubern-Mérida

GEBOREN OP 19 FEBRUARI 1985 TE GIRONA, SPANJE

Promotor: **Prof. dr. ir. N. Karssemeijer**

Copromotoren: **Dr. ir. R. Martí**
Universitat de Girona, Spanje
Dr. ir. B. Platel

Manuscriptcommissie: **Prof. dr. T. Heskes**
Prof. dr. J. Pluim
Technische Universiteit Eindhoven
Prof. Sir. M. Brady
University of Oxford, Verenigd Koninkrijk

The research described in this thesis was carried out at the Diagnostic Image Analysis Group, Radboud University Medical Center (Nijmegen, the Netherlands) and at the Department of Computer Architecture and Technology, University of Girona (Girona, Spain).

This work has received funding from the European Union's Seventh Framework Programme FP7 under grant agreement no 306088 and has been supported by the Spanish Science and Innovation grant nb. TIN2012-37171-C02-01. Albert Gubern-Mérida held a FPU grant AP2009-2835.

Financial support for publication of this thesis was kindly provided by the Faculty of Science, Radboud University Nijmegen.

Automated Analysis of Magnetic Resonance Imaging of the Breast

DOCTORAL THESIS

TO OBTAIN THE DEGREE OF DOCTOR FROM RADBOUD UNIVERSITY NIJMEGEN ON THE AUTHORITY OF THE RECTOR MAGNIFICUS PROF. DR. TH. L. M. ENGELEN, ACCORDING TO THE DECISION OF THE COUNCIL OF DEANS, AND TO OBTAIN THE DEGREE OF DOCTOR FROM UNIVERSITAT DE GIRONA ON THE AUTHORITY OF THE RECTOR DR. S. BONET MARULL, TO BE DEFENDED IN PUBLIC ON FRIDAY, APRIL 24, 2015 AT 12.30 HOURS

BY

Albert Gubern-Mérida

BORN ON FEBRUARY 19, 1985 IN GIRONA (SPAIN)

Supervisor: **Prof. dr. ir. N. Karssemeijer**

Co-supervisors: **Dr. ir. R. Martí**
Universitat de Girona, Spain
Dr. ir. B. Platel

Doctoral Thesis Committee: **Prof. dr. T. Heskes**
Prof. dr. J. Pluim
Technische Universiteit Eindhoven
Prof. Sir. M. Brady
University of Oxford, United Kingdom

The research described in this thesis was carried out at the Diagnostic Image Analysis Group, Radboud University Medical Center (Nijmegen, the Netherlands) and at the Department of Computer Architecture and Technology, University of Girona (Girona, Spain).

This work has received funding from the European Union's Seventh Framework Programme FP7 under grant agreement no 306088 and has been supported by the Spanish Science and Innovation grant nb. TIN2012-37171-C02-01. Albert Gubern-Mérida held a FPU grant AP2009-2835.

Financial support for publication of this thesis was kindly provided by the Faculty of Science, Radboud University Nijmegen.

TABLE OF CONTENTS

1	Introduction	3
1.1	Breast cancer screening	4
1.2	DCE-MRI of the breast	5
1.3	Current clinical use of computer assistance	8
1.4	Thesis outline	9
2	Segmentation of the pectoral muscle in breast MRI	11
2.1	Introduction	13
2.2	Material	14
2.3	Methods	15
2.4	Results	17
2.5	Discussion	19
3	Breast segmentation and density estimation in breast MRI	21
3.1	Introduction	23
3.2	Methods	25
3.3	Experiments and results	31
3.4	Discussion and conclusions	36
4	Validation of volumetric breast density estimations from FFDM using MRI	41
4.1	Introduction	43
4.2	Materials and methods	44
4.3	Results	47
4.4	Discussion	48
5	Automated breast cancer detection in breast MRI	53
5.1	Introduction	55
5.2	Material and methods	56
5.3	Results	65
5.4	Discussion	66
5.5	Conclusion	71
6	Automated detection of breast cancer in prior false-negative screening MRI	73
6.1	Introduction	74
6.2	Materials and methods	75

6.3 Results	78
6.4 Discussion	79
Summary and discussion	81
Samenvatting	89
Resum	93
Publications	99
Bibliography	103
Acknowledgments	121
Curriculum Vitae	125

Introduction

1

1.1 Breast cancer screening

Breast cancer is the most frequently diagnosed cancer among women, worldwide¹. In 2012, 464 000 new cases (13.5% of all cancers) were diagnosed in Europe and 131 000 died from the disease². It is known that survival rate is associated with the stage of the detected cancer. If breast cancer is detected early, chances for survival increase due to more effective treatment, and quality of life of patients decreases less because early detection enables less radical surgery and restrictive use of radiation therapy.

Many countries have introduced breast cancer screening programs to benefit from early breast cancer detection. In regular breast cancer screening programs, asymptomatic women are periodically invited for a mammographic examination. Several studies showed that breast cancer screening by mammography reduces breast cancer mortality in women over age 50 by 25-30%^{3,4}. However, mammography has its limitations as its sensitivity is severely impaired in women with dense breasts. The reason is that dense tissues (fibroglandular and stromal tissues) and breast cancer appear both equally bright on mammographic images as they have the same or similar x-ray attenuation properties. As a result, the risk of missing breast cancer in screening programs while they are still in an early stage of development is increased⁵⁻⁹. In addition, women with dense breasts have a risk of breast cancer four to six times higher than women with no or little dense tissue¹⁰.

Personalized, tailored screening programs have been proposed to improve breast cancer detection in women with dense breasts and other risk factors for developing breast cancer^{11,12}. Those risk factors which are associated with increased risk are family history of breast cancer and the presence of a germ-line mutation of the *BRCA1* or the *BRCA2* gene^{13,14}. In these personalized screening programs, these high risk populations can be screened with a complementary modality, such as ultrasound or Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI). American¹⁵ and European¹⁶ Cancer Societies guidelines recommend the use of DCE-MRI as a complementary imaging modality for women with cumulative lifetime breast cancer risk of more than 20-25%. Other indications for DCE-MRI include, but are not limited to, preoperative staging, evaluation of women treated with neoadjuvant chemotherapy and problem solving in case of inconclusive findings from other modalities.

Compared to mammography, breast DCE-MRI presents higher sensitivity¹⁷⁻²², especially in women with dense breasts²³. However, specificity is more variable, both for screening¹⁶ and characterization²⁴ purposes, since the examination of breast DCE-MRI depends on many factors such as reader expertise and use of adequate visualization techniques. Another limitation of breast MRI is that its analysis requires interpretation of four-dimensional DCE data, as well as correlation to multi-parametric data from

other MRI imaging sequences, and is therefore a time consuming task. Furthermore, recent studies reported potential observer errors such as misinterpretation or oversight of breast cancer lesions that were visible on a follow-up or incident round of screening MRI^{25,26}.

Automated image analysis techniques and Computer-Aided Detection (CAD) systems can be used in order to aid radiologists in reading and interpreting MRI images. This thesis focuses on the investigation of image analysis techniques for the automated interpretation of breast DCE-MRI images.

1.2 DCE-MRI of the breast

An example of an MRI scanner is shown in Fig. 1.1. An MRI system consists of the following components:

- A large magnet to generate the magnetic field and shim coils to make the magnetic field as homogeneous as possible. This magnetic field aligns the hydrogen nuclei of the body parts being imaged.
- A radio-frequency (RF) coil to transmit a radio signal into the body part being imaged. This radio signal is applied after aligning the hydrogen nuclei with the high magnetic field.
- A receiver coil to detect the returning radio signals due to the nuclei relaxation.
- Gradient coils to provide spatial localization of the signals applying additional magnetic fields. These additional magnetic fields can be used to only generate detectable signal from specific locations in the body (spatial excitation) and/or to make magnetization at different spatial locations process at different frequencies, which enables k -space encoding of spatial information.
- A computer to reconstruct the radio signal into the final image (usually by means of Fourier transforms).

The voxel intensity of an MR image is determined by four basic parameters: proton density, T1 relaxation time, T2 relaxation time, and flow. Proton density is the concentration of protons in the tissue in the form of water and macromolecules (proteins, fat, etc.). The T1 and T2 relaxation times define the way the protons revert back to their resting states after the initial RF pulse. The most common effect of flow is loss of signal from rapidly flowing arterial blood.

The signal intensity contrast in an MR image can be manipulated by changing the pulse sequence parameters. A pulse sequence sets the specific number, strength, and



Figure 1.1: Siemens 3T MRI scanner. Image from www.healthcare.siemens.com.

timing of the RF and gradient pulses. The two most important parameters are the repetition time (TR) and the echo time (TE). The TR is the time between consecutive RF pulses. The TE is the time between the initial RF pulse and the echo.

The most common pulse sequences are the T1-weighted and T2-weighted spin-echo sequences. The T1-weighted sequence uses a short TR and a short TE (TR < 1000 msec, TE < 30 msec). The T2-weighted sequence uses a long TR and a long TE (TR > 2000 msec, TE > 80 msec).

Dynamic contrast-enhanced magnetic resonance imaging of the breast is most commonly performed as a combination of several T1-weighted MR image acquisitions over time following contrast agent injection. MRI has several types of contrast agents usually containing paramagnetic metals. The most common type of these agents is the gadolinium chelates. Cancer tends to be a relatively rapid growing tissue needing lots of nutrients and oxygen. This causes cancer cells to stimulate blood vessel growth. The increased vascularity and permeability of the blood vessels causes contrast agent to diffuse out into cancerous tissue more easily than into normal tissue, resulting in higher concentrations and thus higher signal intensity on T1-weighted sequences.

Figure 1.2 shows a graphical representation of the DCE-MRI acquisition process. As stated in the European breast MRI guidelines¹⁶, a dynamic sequence demands at least 3 time points to be measured: one before the administration of the contrast medium, one approximately 2 min later to capture the peak and one in the late phase to evaluate whether a lesion continues to enhance (persistent), shows plateau or shows early washout of the contrast agent. The time course of the signal intensity enhancement over the lesion is of important diagnostic value as demonstrated by Kuhl et al.²⁷.

The optimal number of repetitions is unknown, but 4 or 5 post contrast acquisitions, with data acquisition times in the order of 60 - 120 sec per volume acquisition, are commonly performed in current clinical practice. DCE-MRI of the breast also requires a high spatial resolution with voxel size under 2.5 mm in any direction to be capable of detecting all lesions larger than or equal to 5 mm. During the acquisition process, the patient is placed in prone position. A dedicated bilateral breast coil (see Fig. 1.3) is used. It is also important to mention that, in pre-menopausal women, it is essential to perform breast MRI in the correct phase of the menstrual cycle as enhancing normal breast tissue may otherwise complicate the interpretation of the study^{25,28}.

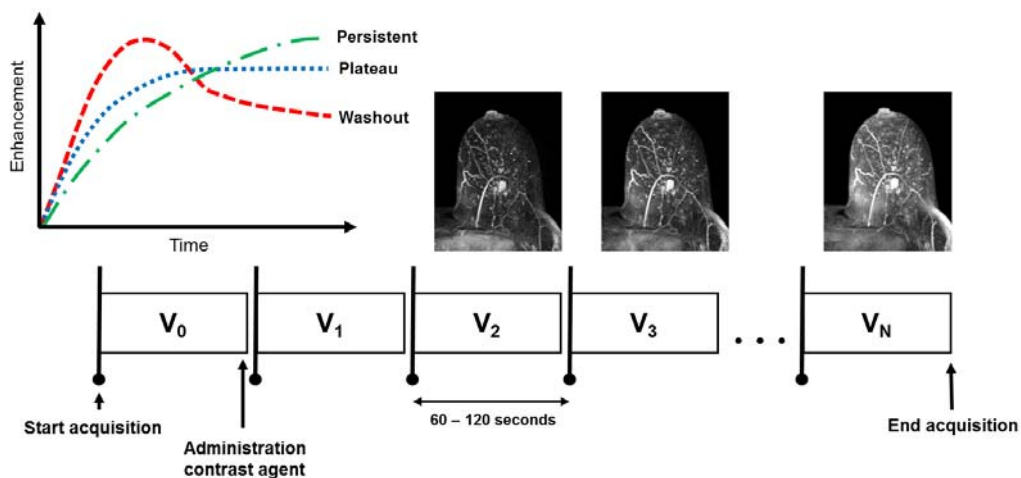


Figure 1.2: Breast DCE-MRI acquisition scheme. Several volumes are acquired before (V_0) and after (V_{1-N}) contrast agent injection. Signal enhancement computed over the region of the lesion can be classified as persistent (green), plateau (blue) and early washout (red).

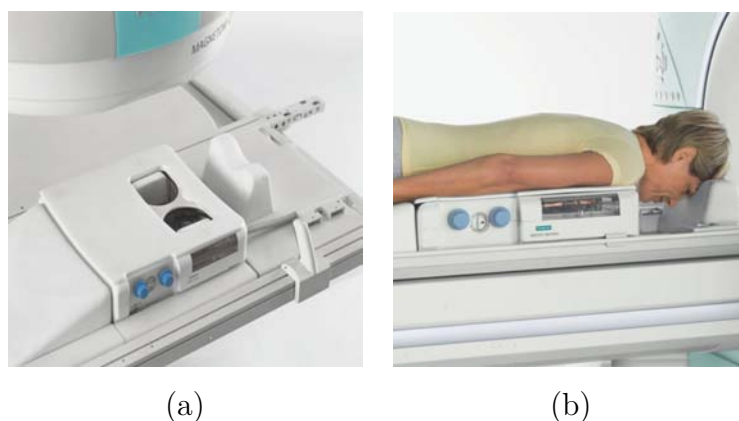


Figure 1.3: (a) Dedicated bilateral breast coil. (b) Patient is placed in prone position with both breasts hanging in the coil loops. Images from www.healthcare.siemens.com.

For diagnostic interpretation of breast lesions, T1-weighted images and subtracted images and their maximum intensity projection for all the available times points are

analyzed (see Fig. 1.4). Morphological and kinetic information (curve type) of the breast lesions are assessed according to the ACR Breast Imaging and Data System (BI-RADS) breast MRI lexicon²⁹.

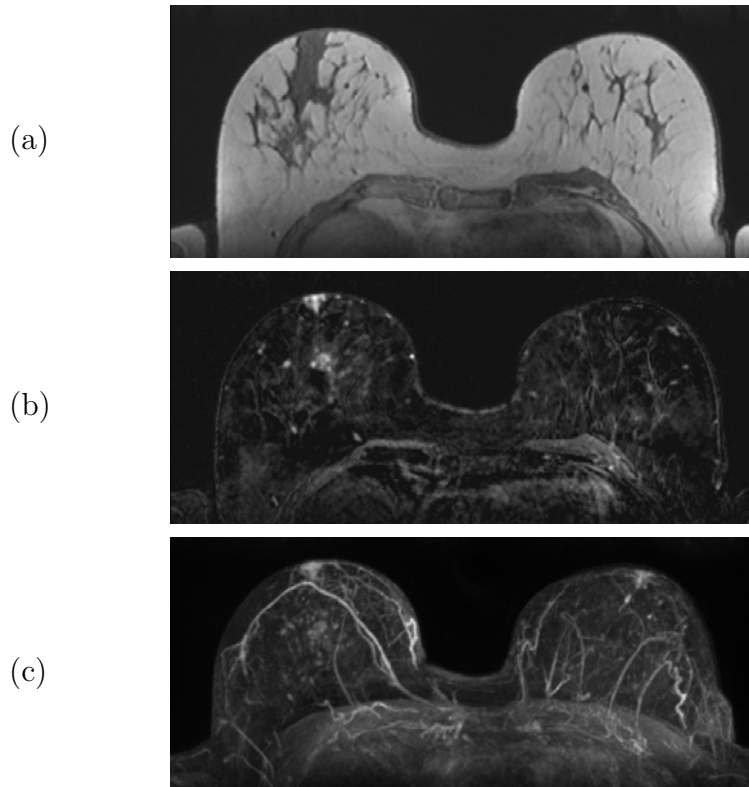


Figure 1.4: (a) Axial slice of the pre-contrast T1-weighted volume and (b) corresponding axial slice extracted from the subtraction volume computed at the first post contrast time point. Maximum intensity projection image of the subtracted volume is shown in (c).

1.3 Current clinical use of computer assistance

In current clinical practice, dedicated workstations are used to assist the radiologist in detection and classification of breast lesions in DCE-MRI data. These systems visualize a single DCE-MRI study by displaying T1-weighted, subtracted and maximum intensity projection images of all the available times points of a desired region in addition to the kinetic curve. Moreover, these dedicated workstations provide an automated kinetic assessment by color-coding the intensity changes per voxel during enhancement of the breast tissue. The visualization of the images combined with this automated assessment aid in the interpretation of patterns of contrast enhancement (persistent, plateau and washout enhancement) across a series of MRI volumes³⁰. However, human interaction is still required to identify and characterize suspicious areas in four-dimensional data. Therefore, interpretation of breast DCE-MRI is still a time consuming task. Human

interaction increases the risk of misinterpreting or overlooking breast lesions^{25,26} and may cause inter- and intra-observer variability³¹.

Due to these limitations, there is still the need to develop automated image analysis tools and computer-aided detection systems to aid in the interpretation of DCE-MRI of the breast. These automated techniques have the potential to decrease the workload of radiologists and improve the diagnosis. In screening mammography, computer-aided detection systems have been considered as a tool to relieve the shortage of trained readers for double reading^{32,33}. These systems can also be used as an aid to help the first reader localizing the abnormality more quickly and more accurately, suppressing the chance that lesions are overlooked by radiologists due to fatigue or lack of experience. Different studies³⁴⁻³⁷ have shown that use of CAD can improve reader performance of breast cancer detection or screening. In the USA, the majority of mammograms are nowadays read with CAD support.

CAD systems for breast DCE-MRI are being developed³⁸⁻⁴⁰. These CAD systems follow a similar pipeline composed of motion artifacts correction, breast segmentation, voxel candidate detection and regional candidate detection analysis. However, they are still in an initial stage of development and, therefore, further research is required to improve each of these automated image analysis steps that are not easy tasks on their own. For instance, some authors focused on automated segmentation of the breast^{41,42}, which requires the delineation of the chest wall and the boundary between breast and background (or air). This is a difficult task due to the large shape variations across patients. Another complicating factor that stands in the way of any image processing application is the MR signal intensity variability from patient to patient due to differences in size and anatomy of each patient, the use of different acquisition parameters and/or the effect of image inhomogeneities.

The development of automated image analysis techniques is also important for their use in other applications that can improve the clinical outcome in breast MRI. Automatic linkage of current and prior examinations⁴³, breast density estimation in breast MRI⁴⁴ and automated breast MRI quality assessment are some examples.

1.4 Thesis outline

This thesis describes methods for quantitative analysis of magnetic resonance imaging of the breast. The main focus is breast density estimation, as dense tissue has been related to the risk of breast cancer development, and automated lesion detection to aid in the interpretation of DCE-MRI of the breast. Other techniques have also been developed to serve the main goal including image normalization, pectoral muscle segmentation and breast segmentation, among others.

The outline of this thesis is as follows. Chapter 2 provides two automated pectoral muscle segmentation methods. The delineation of the pectoral muscle is essential to segment the breast. In particular, we investigated the performance and complexity of two novel atlas-based methods for pectoral muscle segmentation in breast MRI. Chapter 3 presents a fully automatic approach to segment the breast and estimate breast density. Other image analysis techniques such as image normalization are also described. In chapter 4, a method for measuring volumetric breast tissue estimates from digital mammograms^{45,46} is validated by comparing its results to volume estimates that were obtained from breast MRI data. The volumetric estimates from MRI were obtained using an improved version of the method described in chapter 3. Chapter 5 describes a CAD system for the automatic detection of breast cancer. In chapter 6, the potential of the CAD system to detect breast cancer which was overlooked or misinterpreted by a radiologist in a breast screening program with MRI is investigated. The thesis will be finalized with a summary and discussion chapter.

Segmentation of the pectoral muscle in breast MRI

2

Albert Gubern-Mérida, Michiel Kallenberg, Robert Martí, Nico Karssemeijer

Original title: Segmentation of the pectoral muscle in breast MRI using atlas-based approaches

Published in: Lecture notes in computer science (2012); 7511(1):371–378

Abstract

Pectoral muscle segmentation is an important step in automatic breast image analysis methods and crucial for multi-modal image registration. In breast MRI, accurate delineation of the pectoral is important for volumetric breast density estimation and for pharmacokinetic analysis of dynamic contrast enhancement. In this paper we propose and study the performance of atlas-based segmentation methods evaluating two fully automatic breast MRI dedicated strategies on a set of 27 manually segmented MR volumes. One uses a probabilistic model and the other is a multi-atlas registration based approach. The multi-atlas approach performed slightly better, with an average Dice coefficient (DSC) of 0.74, while with the much faster probabilistic method a DSC of 0.72 was obtained.

2.1 Introduction

Automatic identification of pectoral muscle is an important step in methods for automatic breast cancer assessment in most image modalities. For instance, in mammography, the most used image modality in screening programs, the detection and removal of the pectoral muscle is often used to remove false positive marks of Computer Aided Detection (CAD) systems⁴⁷. In Magnetic Resonance Imaging (MRI) of the breast, the image modality employed in the presented work, the importance of the pectoral muscle detection has recently been recognized in two applications. Firstly, breast tissue density has been identified as an important risk factor for developing breast cancer, being four times larger in women with a breast density higher than 75%, compared to those with little or no density⁷. Breast MRI provides a good tissue contrast between fibroglandular and fatty tissues and a three-dimensional characterization of breast composition. These good properties in the breast tissue have been a strong reason to use breast MRI in breast density measurement^{48–50}. However, the contrast between pectoral muscle and dense tissue is poor. Hence, a first step to separate the breast from the body is commonly essential. This separation is not trivial due to the large shape and intensity variations in the pectoral muscle of different patients. Some solutions are present in the literature: boundary tracing or spline fitting without⁵¹ and with manual intervention^{48,49} and delineation of the whole breast using breast models⁵², but none of them completely delineates the pectoral muscle.

Secondly, MRI is often used with a contrast agent for lesion detection. For a better interpretation of contrast enhancement lesions, researchers have tried to incorporate pharmacokinetic modeling to the interpretation of the MRI. Some of these models require calibrations with respect to reference tissues and make use of the signal intensity of specific regions for determining physiological measures⁵³. In breast MRI, the pectoral muscle can be used as a reference tissue given its properties.

Atlas-based segmentation has been shown to be a powerful technique for automatic delineation of anatomical structures in different 3D image modalities^{54,55}. Multi-atlas and probabilistic approaches are the most commonly used strategies. By definition, the former is supposed to obtain more precise segmentations than the latter. However, multi-atlas approach is far more time consuming. There has been only one initial attempt that uses an atlas strategy for breast MRI segmentation⁵⁰, but the segmentation of the pectoral muscle was not the main interest of the work. Moreover, the method followed a probabilistic approach using one reference atlas, which could have some limitations. As shapes are highly variable, the reference choice affects final results.

The novelty of this paper consists in the study of fully automatic atlas-based methods for pectoral muscle segmentation in breast MRI in terms of performance and com-

plexity. A dedicated multi-atlas approach based on Klein et al.⁵⁵ is proposed (see section 2.3.3) and compared to the probabilistic approach of Gubern-Mérida et al.⁵⁰ (see section 2.3.2). An original breast MRI registration framework focused on the body has been also defined and used in both methods (see section 2.3.1). To our knowledge, no similar studies are found in the literature. Advantages and inconveniences of both strategies are discussed in sections 2.4 and 2.5 and a solution to obtain a reasonable time-accuracy trade-off is proposed.

2.2 Material

The data set used to evaluate the segmentation results and build the atlases consists of 27 pre-contrast T1-weighted MR breast scans obtained from different patients. Breast MRI examinations were performed on a 1.5 T system (Siemens 1.5T, Magnetom Vision), with a dedicated breast coil (CP Breast Array, Siemens, Erlangen). The pixel spacing differed between volumes with values ranging from 0.625 mm to 0.722 mm. The slice thickness was 1.3 mm and the volume size was 512 x 256 x 120 voxels. Patients were scanned in prone position.

Three experienced observers performed manual segmentations. Two of them manually segmented only the pectoral muscles of 8 cases. The third one manually segmented each of the 27 MR volumes into 7 classes: background, fatty tissue, glandular tissue, pectoral muscles, lung area and the heart. The seventh class is the "other" class and refers the previous non-labeled voxels of the thorax. Annotations were done every 5-10 slices and linear interpolation was applied to obtain the complete labeling. When needed, and especially for heart, lungs and pectoral muscles, accurate manual delineation was performed with a smaller slice interval. For the manual segmentation of background, fatty and fibroglandular tissue, thresholding was applied over regions of interest provided by the reader. Fig. 2.1 shows an example of a MRI slice on an axial view and the manual delineation of the mentioned classes. One should note the complexity of performing such ground truth annotations, where each volume takes approximately 45 minutes in a dedicated breast MRI annotation environment.

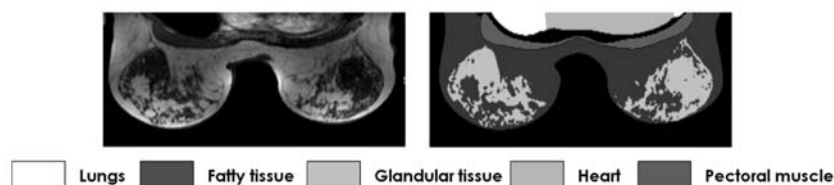


Figure 2.1: Breast MR scan on an axial slice with the manual annotation of the different structures.

2.3 Methods

Atlas-based strategies are characterized by the use of pre-labeled images, usually manually obtained, to perform the automatic segmentation of new images, also called targets. They employ registration algorithms, which play an important role for the final segmentation. Section 2.3.1 explains the mapping algorithm used by the two atlas-based approaches evaluated in this paper. Section 2.3.2 briefly describes the construction and the use of a probabilistic atlas in a Bayesian framework segmentation⁵⁰. Finally, in section 2.3.3 we report the proposed multi-atlas segmentation algorithm for the delineation of the pectoral muscle in breast MRI.

2.3.1 Registration

Registration is an important step in atlas-based segmentation algorithms. Without an accurate transformation between the structures we aim to segment, the segmentation can not perform accurately. For this reason we developed a registration framework focused on the body area. We observed that the sternum is always localized between pectoral muscles. Hence, by accurately localizing the sternum the pectoral muscles can be aligned. Our registration approach is initialized by detecting the sternum of the subjects. Automatic sternum detection is described in Sec. 3.2.1 of this thesis. Then, the volumes are cropped at 2 cm distance anterior to the sternum position to focus the registration on the area of the body. By doing so, most breast tissue is removed and can not negatively bias the final mapping of body structures. The 2 cm distance anterior to the sternum ensures that pectoral muscle voxels are not discarded.

The registration process is composed by two stages. First, a translation transform is performed, where translation along the y axis is defined by the distance between y -coordinates of both sternums. Translation along x and z axis is found by optimizing the similarity measure. The second stage is a non-rigid transform based on B-Splines registration in a multi-resolution scheme using a stochastic gradient descent optimizer. Three resolutions were defined. B-Splines grid spacing was set to 32, 16 and 8 mm for each of the resolutions taking the size of the pectoral muscle into account. The similarity measure maximized by the whole framework was normalized cross correlation (NCC) as all the datasets were acquired with the same modality. Elastix⁵⁶ was used for the implementation.

2.3.2 Method 1: Probabilistic atlas-based segmentation

In the presented atlas-based segmentation method, a probabilistic atlas is used in a Bayesian framework to provide an accurate probability distribution for the pectoral and the thoracic area. Following a leave-one-out evaluation strategy, for the segmentation of

each patient, a full probabilistic atlas was built offline with the 26 remaining patients. These 26 patients and their segmentations were mapped using the registration method explained previously into the same reference space. The probabilistic atlas was created by computing the frequency with which each location was labeled as a specific organ. A common reference space was used for all the experiments by visually selecting an extra patient which has normal appearance. The reference case, or anatomical image of the atlas, was not included in the evaluation set.

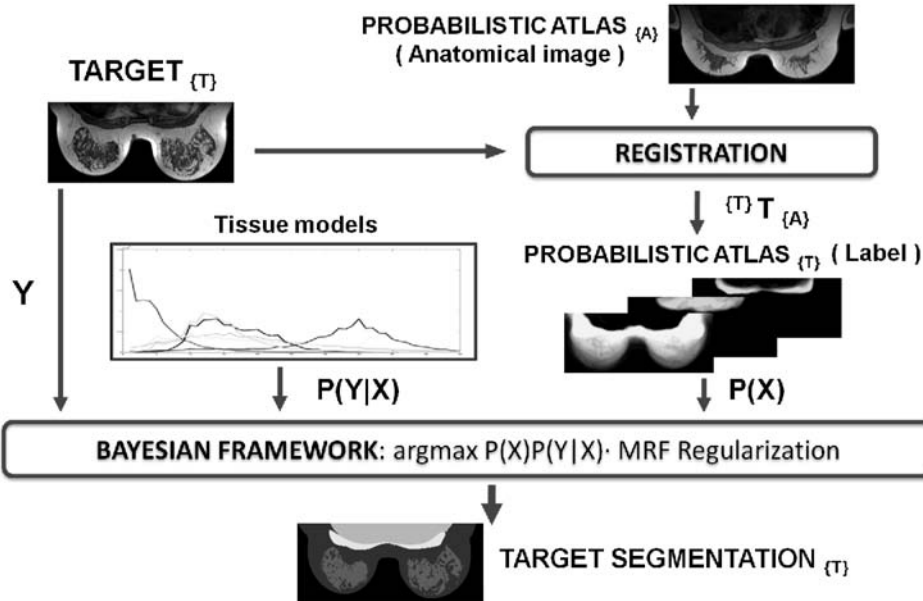


Figure 2.2: Probabilistic atlas segmentation approach overview.

Figure 2.2 shows the general schema of the segmentation framework with Bayesian voxel classification algorithm incorporating the use of the probabilistic atlas. From top to bottom, the probabilities of the atlas are mapped by registration of section 2.3.1 onto target image space $\{T\}$ using the anatomical image of the atlas. The probabilistic atlas, the tissue models (previously built from the scans and manual segmentations of the data set) and the target are provided to the Bayesian framework as a prior probability $P(X)$, conditional probability $P(Y|X)$ and set of intensity values Y , respectively. The Bayesian framework estimates the segmentation X that maximizes $P(X)P(Y|X)$ and also includes a Markov Random Field (MRF) regularization to smooth the segmentation taking neighborhood information into account⁵⁰.

2.3.3 Method 2: Multi-atlas segmentation

Multi-atlas segmentation approaches consider all the volumes of the dataset and their manual segmentations as individual atlases. The term atlas is defined as the pair of the anatomical image (MRI volume) and its manual segmentation or label. The process to obtain an automatic segmentation for a target volume is illustrated in figure 2.3. First,

given the target volume T , all the atlases are mapped onto the target space using the registration algorithm of section 2.3.1. Subsequently, the deformed anatomical images are compared to the target to perform a selection of the most similar atlases. The selection is based on the Normalized Cross Correlation similarity measure and a ratio defined as follows:

$$r_i = \frac{NCC(T, A_i \circ M_i)}{\max_j NCC(T, A_j \circ M_j)}, \quad (2.1)$$

where M refers to the mapping between the target and an atlas and j refers to the deformed atlas with maximum similarity. An atlas A_i is selected if it satisfies $r_i \geq \varphi$. A value of $\varphi = 0.9$ empirically appeared to be the best value for our results.

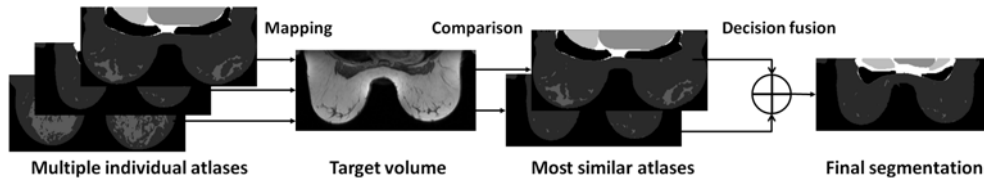


Figure 2.3: Multi-atlas segmentation approach overview.

Finally, the selected deformed atlas labels are fused to yield a single final segmentation of the patient or target image. This step is called decision fusion and defines how the deformed segmentations of the selected atlas are combined. In this work we have made use of majority voting method, which was proven to give good results in⁵⁵.

2.4 Results

In a leave-one-out experiment we evaluated the probabilistic and the multi-atlas segmentation frameworks on 27 patients. Each segmented case was not included for the construction of the probabilistic atlas or within the set of individual atlases respectively. The quality of the segmentation was measured by determining the similarity of the segmentation with the ground truth using the Dice Similarity Coefficient (DSC). DSC was chosen as it is commonly used in the literature^{50,55}. For all cases we manually discarded initial and last slices which do not contain relevant information or are clearly affected by noise. Figure 2.4(a) shows a box plot with DSC values for each method. Segmentation results are similar (DSC median of 0.76 for both and DSC mean \pm sd of 0.72 ± 0.09 and 0.74 ± 0.06 for probabilistic and multi-atlas respectively), but multi-atlas framework slightly outperforms the probabilistic. These results can be better seen in figure 2.4(b), where DSC values of each case using both methods are shown.

Lower DSC values are mainly due to the registration process not being able to compensate the differences between volumes. This is more the case of the probabilistic

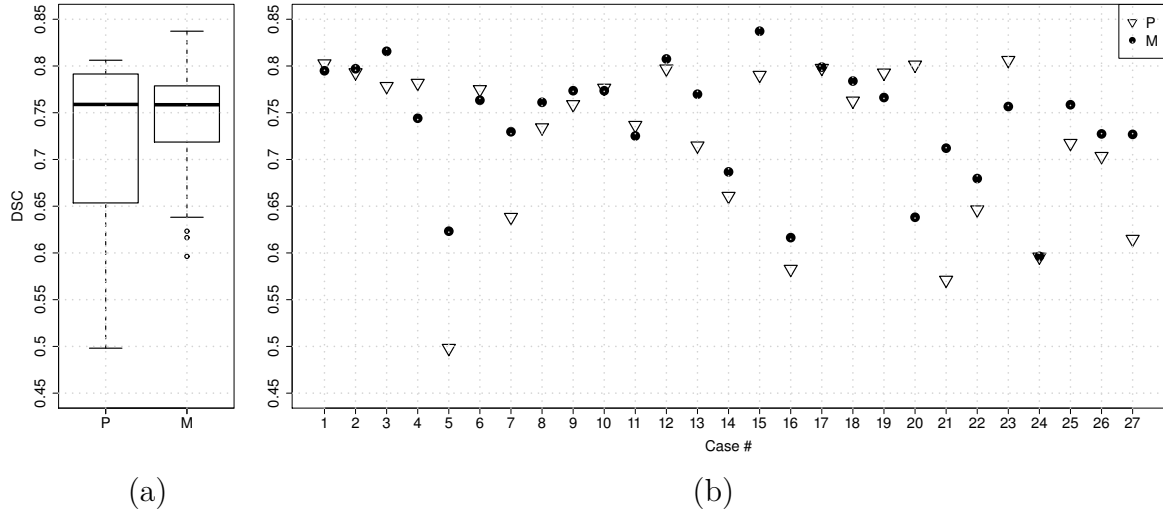


Figure 2.4: (a) Box plot with segmentation DSC for pectoral segmentation and (b) DSC segmentation results for each of the 27 cases using probabilistic (P) and multi-atlas (M) approaches.

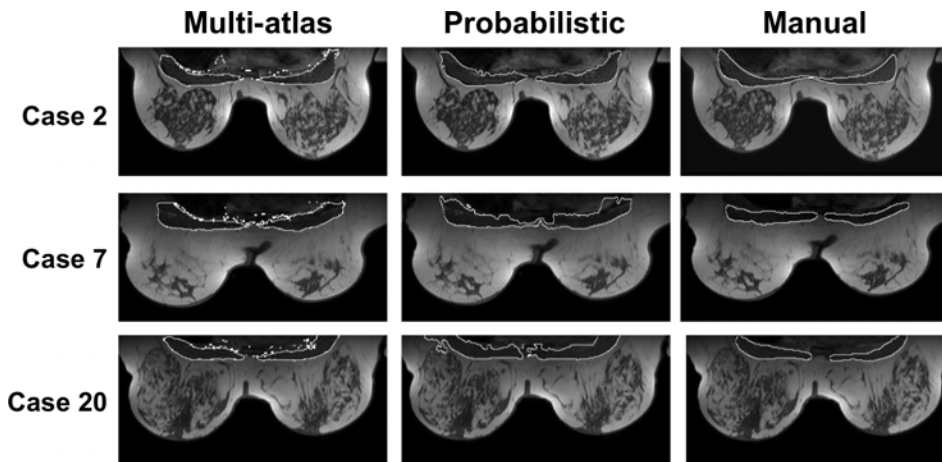


Figure 2.5: Intermediate slices from 3 different cases and their segmentations.

approach as the method uses only one registration with a single reference. In those cases (see case 7 for instance), the multi-atlas approach performs better as it includes multiple registrations and selects the best ones. Only in one case (number 20) the probabilistic approach obtains results much better than multi-atlas, where pectoral muscle segmentation in initial slices is not really precise (labeled as thorax instead). However, accurate delineations in intermediate slices are obtained for both methods as it is illustrated in figure 2.5, where three examples of automatic and manual segmentations are shown.

Finally, since no previous works performed pectoral segmentation in breast MRI, inter-observer variability was computed by 3 viewers over 8 manual segmentations. DSC mean of 0.70 ± 0.12 and median of 0.72 were obtained, lower than the DSC values achieved by the automatic atlas-based approaches.

All the tests have been launched on Intel(R) Core(TM)2 Quad CPU Q9550 2.83GHz.

Starting with the common step, registration between two volumes takes $t_r \approx 12$ min. The complexity time for multi-atlas segmentation is explained as $t_{multi-atlas} \approx N \times (t_r + t_a) + n \times t_f$, where N is the number of individual atlases ($N = 26$), t_a the time to compute the mapping and the comparison of an individual anatomical atlas ($t_a \approx 4$ min.), n the number of selected atlases and t_f the time to propagate and fuse an atlas labeled image ($t_f \approx 3$ min.). In the best scenario, being only one atlas selected ($n = 1$), the computation time to obtain a segmentation using a multi-atlas approach is $t_{multi-atlas} \approx 419$ minutes (7 hours). The complexity time for the probabilistic approach is defined as $t_{probabilistic} \approx t_r + t_p + t_b$, where t_p is the time to map the probability distributions to the target space ($t_p \approx 8$ min) and t_b the time to perform the segmentation based on Bayesian theory ($t_b \approx 10$ min). Approximately, $t_{probabilistic} \approx 30$ min.

2.5 Discussion

In this work, the atlas-based methodology has been studied to perform the complete delineation of the pectoral muscle in breast MRI, which has not been done previously. Fully automatic and dedicated multi-atlas and probabilistic frameworks have been proposed and tested on 27 different patients.

The obtained results are satisfactory in both frameworks, with DSC values higher than the computed inter-observer variability. It proves the high reliability of atlas-based segmentation methods to perform pectoral delineations. However, we are aware that the evaluation and the construction of the atlases were performed with annotations from a single viewer, as obtaining 3-dimensional manual segmentations is a time consuming task. The low inter-observer DSC value explains the difficulty and subjectivity to delineate the pectoral muscle. Its shape has high-variability and cartilage, intercostal muscles and fatty tissue also appear in the area. The inclusion of these tissues depends on the observer opinion.

As it was expected, multi-atlas segmentation appears to be more consistent than the probabilistic. This is explained by the fact that the multi-atlas approach includes an atlas selection step to choose the most similar atlas compared to the segmented volume. In the probabilistic framework, when the target differs considerably from the reference atlas and the registration can not compensate the differences, the final segmentation becomes affected with slightly poorer results. However, the computation time for a multi-atlas segmentation is 14 times larger.

Considering the influence of atlas selection, in future work we will study a multi-probabilistic atlas framework. A larger dataset will be created with annotations from different observers. We will group different breast MRI populations based on shape.

For each population, a probabilistic atlas will be built. The most similar atlas to the image at hand will be chosen for segmentation.

Breast segmentation and density estimation in breast MRI

3

Albert Gubern-Mérida, Michiel Kallenberg, Ritse M. Mann, Robert Martí, Nico Karssemeijer

Original title: Breast segmentation and density estimation in breast MRI: a fully automatic framework

Published in: IEEE Journal of Biomedical and Health Informatics (2015);
19(1):349–357

Abstract

Breast density measurement is an important aspect in breast cancer diagnosis as dense tissue has been related to the risk of breast cancer development. The purpose of this study is to develop a method to automatically compute breast density in breast Magnetic Resonance Imaging (MRI). The framework is a combination of image processing techniques to segment breast and fibroglandular tissue. Intra- and inter-patient signal intensity variability is initially corrected. The breast is segmented by automatically detecting body-breast and air-breast surfaces. Subsequently, fibroglandular tissue is segmented in the breast area using Expectation-Maximization. A data set of 50 cases with manual segmentations was used for evaluation. Dice Similarity Coefficient (DSC), total overlap, False Negative Fraction (FNF) and False Positive Fraction (FPF) are used to report similarity between automatic and manual segmentations. For breast segmentation, the proposed approach obtained DSC, total overlap, FNF and FPF values of 0.94, 0.96, 0.04 and 0.07, respectively. For fibroglandular tissue segmentation, we obtained DSC, total overlap, FNF and FPF values of 0.80, 0.85, 0.15 and 0.22, respectively. The method is relevant for researchers investigating breast density as a risk factor for breast cancer and all the described steps can be also applied in computer aided diagnosis systems.

3.1 Introduction

Breast Magnetic Resonance Imaging (MRI) is a technique employed for (1) screening for breast cancer in high risk patients, (2) evaluation of tumor extent in specific groups of patients with breast cancer (e.g. patients with invasive lobular carcinoma), (3) evaluation of tumor response to chemotherapy treatment and, (4) trouble shooting in case of inconclusive findings from other modalities¹⁶. It provides good tissue contrast between fibroglandular (also referred to as dense) and fatty tissues, thus allowing three-dimensional characterization of breast composition. Segmentation of the different structures visible in a breast MRI is needed to perform an automatic analysis of such images. Some examples of applications in breast MRI which require an initial segmentation step are multi-modal breast image registration, computer aided analysis of dynamic contrast enhanced MRI⁵⁷, and breast density assessment^{7,58}. Related to the latter, breast density has been identified as an important risk factor for developing breast cancer, with risk being four times larger in women with a breast density higher than 75%, compared to those with little or no density^{7,58}. Automated segmentation of breast density in breast MRI is the main focus of this work.

In general two steps are required to obtain 3D breast density measurement from MRI: Breast segmentation and fibroglandular tissue segmentation. The segmentation of the breast is initially performed to exclude other tissue that does not belong to the breast, such as pectoral muscle. The separation of the breast from the body is a difficult task. Complicating factors are the large shape variations of pectoral muscles across different patients and the similarity between intensity distributions of the MRI signal in muscle and fibroglandular tissues. Other issues are caused by the lack of agreement on the anatomical extent of the breast. Because no automated segmentation method was available, manual intervention has been mostly required^{44,49,59-61} in the literature. In some studies the definition of the breast-body interface was made by manually defining a straight line^{59,60}. Other studies combined edge detection filters and manual outlining to delineate the breast volume^{49,61,62}. Nie et al.⁴⁴ used an approach for breast volume segmentation that also requires some manual intervention when the chest wall is connected to fibroglandular tissue. Their algorithm starts with the detection of body landmarks and is followed by Fuzzy C-Means (FCM), B-spline fitting and dynamic searching. However, the thoracic spine, which is used as a body landmark, is not always visible in all acquisition orientations and field of views. Fully automatic breast segmentation has been addressed by a few authors. Wang et al.⁴¹ used second derivative information represented by the Hessian matrix to delineate chest wall and air-breast boundary. Another example is found in the work of Gallego-Ortiz et al.⁴², where a method based on 3D edge detection combined with probabilistic atlas of the

breast is proposed to extract the breast. Both approaches led to satisfactory breast segmentation results, but segmentation of internal structures in the breast, such as fibroglandular tissue, was not included.

Regarding the second step, different techniques have been applied to segment the fibroglandular tissue in breast MRI. Interactive thresholding of the breast tissue signal intensity was used to study the correlation between MRI percent density and area-based^{59,62} and volume-based⁶⁰ mammography percent density. Alternatively, other studies proposed a two-compartmental model⁶¹, which linearly combines pure-tissue signals, and FCM algorithm^{44,49,63} to segment fatty and fibroglandular tissue. More recently, an atlas-aided probabilistic model-based method has been also described⁶⁴, where breast tissue is characterized by a mixture model and a fibroglandular tissue probabilistic atlas is used as a prior likelihood in combination with FCM. Among all these studies on breast density estimation on MRI, only Wu et al.^{63,64} presented two fully automatic fibroglandular segmentation approaches integrated with automated breast segmentation. In both approaches, the outline of the breast is initially defined on each slice of the MR volume using context information. In one approach⁶³, the segmentation of fibroglandular tissue is performed using 2D FCM-alone. In the other method⁶⁴, prior learned fibroglandular tissue likelihood is incorporated to FCM to segment fibroglandular tissue. A weak point of these fully automatic approaches is that the employed breast and fibroglandular tissue segmentation algorithms are 2D segmentation methods which are applied to MR volumes on a slice-by-slice basis.

In this work, we present a framework that takes into account the 3-dimensionality nature of breast MRI to automatically segment breast and fibroglandular tissue. In combination, these can be used for automated breast density estimation, which requires computation of the relative amount of fibroglandular tissue in the breast. The framework consists of different image processing steps. Firstly, signal intensity inhomogeneities are corrected. Subsequently, a 3D probabilistic atlas-based approach⁶⁵, previously used for pectoral muscle segmentation, is applied to segment the breast. The Expectation-Maximization (EM) algorithm⁶⁶ is finally used to estimate the image intensity distributions of breast tissue and automatically discriminate between fatty and fibroglandular tissue. The novel contribution of the paper is to extend the work presented by Gubern-Mérida et al.⁶⁵ for breast segmentation and breast density estimation in breast MRI. Breast segmentation step is evaluated with 27 MRI cases fully annotated by an experienced observer. The fibroglandular tissue segmentation step, the final aim of this work, is evaluated with 50 MRI cases compared to the annotations provided by 4 different observers.

3.2 Methods

A general overview of the process to estimate the breast density segmentation is shown in Fig. 3.1. Three preprocessing algorithms are initially applied: firstly, image inhomogeneities are corrected using the N3 bias field correction algorithm⁶⁷ to correct signal intensity variations within the same structure of one specific case. The N3 is a non-parametric method which was designed to be applied on early stages of automated data analysis and does not require a model of the tissue intensities. Secondly, the sternum is detected, which is used as an important landmark in different parts of our algorithm. Thirdly, intensities of the MR images are normalized to compensate for inter-patient signal intensity variability.

The segmentation starts separating the body from the breast. The body consists of lungs, heart, pectoral muscle, thoracic area and fat outside the base of the breast. The breast is connected to the pectoral muscle and is composed by fatty and dense tissues⁶⁸. A probabilistic atlas, which contains spatial information of pectoral muscle, lungs, heart, thorax and breast tissue, is used to exclude the body from the breast⁶⁵. Finally, the breast volume is defined and the dense tissue is delineated using the EM algorithm. Each step is sequentially presented in the following sections.

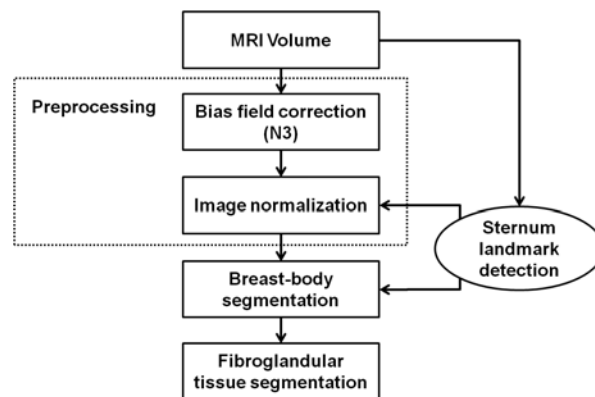


Figure 3.1: General overview of the process for dense tissue segmentation in breast MRI.

3.2.1 Sternum landmark detection

The sternum is a bone localized between pectoral muscles and is always visible in centered axial slices. The use of a breast coil ensures that the sternum is roughly in the center of the image. We consider the sternum as an important landmark to guide the separation between body and breast (see Fig. 3.2(a)) and we use the y -coordinate of the outer edge of the sternum in two stages of our method.

To detect the outer edge of the sternum, we observed that the pectoral muscle border appears as the surface with the strongest edges in the image with maximum

positive gradient in y direction, using the coordinate system defined in Fig. 3.2. Taking advantage of this observation we implemented a first-derivative-based filter to enhance this edge and detect the sternum landmark point. Note that throughout this paper a voxel position is noted as a 3-dimensional vector \mathbf{r} with components r_x , r_y and r_z . Given an MR image $I(\mathbf{r})$, the first derivative volumes $I'_x(\mathbf{r}) = \frac{\delta}{\delta_x}G_{\sigma_1}(I(\mathbf{r}))$, $I'_y(\mathbf{r}) = \frac{\delta}{\delta_y}G_{\sigma_2}(I(\mathbf{r}))$, $I'_z(\mathbf{r}) = \frac{\delta}{\delta_z}G_{\sigma_3}(I(\mathbf{r}))$ are computed in each direction at different Gaussian scales: $\sigma_1 = \sigma_3 = 5$ mm and $\sigma_2 = 1$ mm. The output of the filter is defined as

$$F(\mathbf{r}) = \begin{cases} I'_y(\mathbf{r}) - (I'_z(\mathbf{r}) + I'_x(\mathbf{r})) & \text{if } I'_y(\mathbf{r}) \geq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (3.1)$$

$F(\mathbf{r})$ can be interpreted as the likelihood of a specific voxel at location \mathbf{r} to belong to an edge with y -gradient. The expression $(I'_z(\mathbf{r}) + I'_x(\mathbf{r}))$, subtracted from $I'_y(\mathbf{r})$, can be seen as a penalization term because high gradient values in z and x directions are not expected in voxels that belong to the pectoral muscle surface. Figure 3.2(b) shows an example.

The output of $F(\mathbf{r})$ is binarized by an adaptive threshold set to $\max(F(\mathbf{r}))/4$. We identify the distinct structures of the binary volume by determining the connected components⁶⁹. The largest structure is identified as the pectoral boundary and the rest is excluded. By doing so, smaller edges that do not belong to the pectoral boundary, such as skin folds, are removed. Finally, a box-shaped Volume of Interest (VOI) is placed in the center of the volume (see dotted box of Fig. 3.2(b)). The sternum landmark is determined by computing the center of mass of the voxels of the largest structure which are included in the VOI. In our experiments, the size of the VOI was set to 34 x 156 x 14 mm to ensure that the sternum landmark was detected at any y location of the center of the volume.

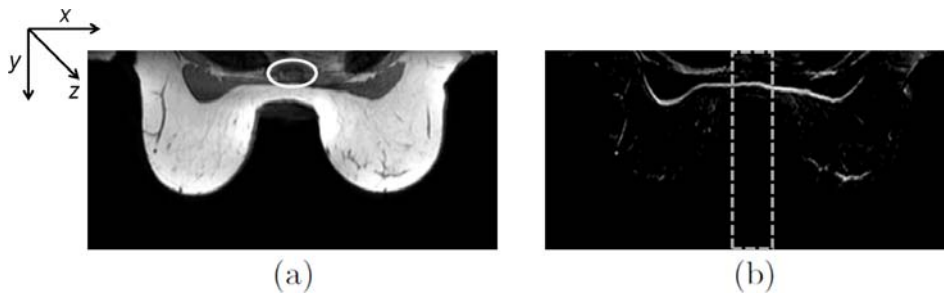


Figure 3.2: (a) MRI axial slice, with sternum indicated with a circle, and (b) the output obtained applying the filter to detect the sternum point with a VOI (dotted box), as described in Eq. 3.1.

3.2.2 Image normalization

Bias field correction minimizes the signal intensity variability of a given tissue within a single acquisition. However, voxel values belonging to a specific structure vary from

patient to patient due to differences on the build of each patient and differences on the acquisition parameters. Ignoring the inter-patient variability can severely hamper the final segmentation of the different tissues, especially in steps where image intensity is used as a feature. Hence, signal intensity of all the scans needs to be standardized for segmentation.

Among others, the sensitivity of the coil is one of the most important factors that cause the difference between signal intensities of different scans. Thus, if we define S_t as the MR intensity corrected for coil sensitivity, we can write $I(\mathbf{r}|x_{\mathbf{r}} = t) = \gamma S_t$, where $x_{\mathbf{r}}$ is the label at location \mathbf{r} , t is the tissue type and γ is the factor that generates the difference.

The factor γ can be canceled out by using the mean intensity value of a reference tissue of the same scan:

$$\hat{I}(\mathbf{r}|x_{\mathbf{r}} = t) = \hat{I}_t = \frac{\gamma S_t}{\gamma S_{ref}} = \frac{I_t}{I_{ref}}, \quad (3.2)$$

where \hat{I}_t is the normalized image intensity value for a tissue t and I_{ref} is the image intensity value of the reference tissue of the same scan. The fatty tissue was selected as the reference as it is the most prominent tissue of the image and its intensity distribution is easier to estimate than the distribution of other structures.

The estimation of the mean fatty tissue value, I_{ref} in Eq. 3.2, is performed in a VOI containing all the voxels of the MR scan which are located anterior to the sternum. Sternum detection is described in section 3.2.1. The dotted box of Fig. 3.3 illustrates the VOI, which contains voxels belonging to fatty and glandular tissues, background and, in some cases, voxels which belong to the pectoral muscles. The background is excluded using Otsu thresholding⁷⁰, leaving fatty, dense and pectoral muscle tissues. The fatty mean intensity value is computed by means of the Expectation-Maximization algorithm⁶⁶. Pectoral muscle and fibroglandular tissue have similar voxel values and can be considered as a single class. The image intensity distributions of the two classes (fatty and other tissues) are estimated by fitting two Gaussian distributions.

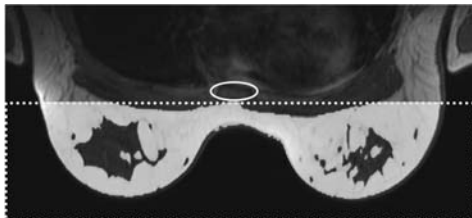


Figure 3.3: Axial breast MR slice: the sternum is localized (ellipse) to define the VOI to estimate the mean value of fatty tissue. Dotted box shows the VOI. The white highlighted tissue within the VOI corresponds to the fatty tissue used as reference tissue in the image normalization process.

As an illustration, Fig. 3.4 shows the MR signal intensity distributions of the whole

image of 5 different cases before (a) and after (b) normalization. A better overlap can be observed after normalization.

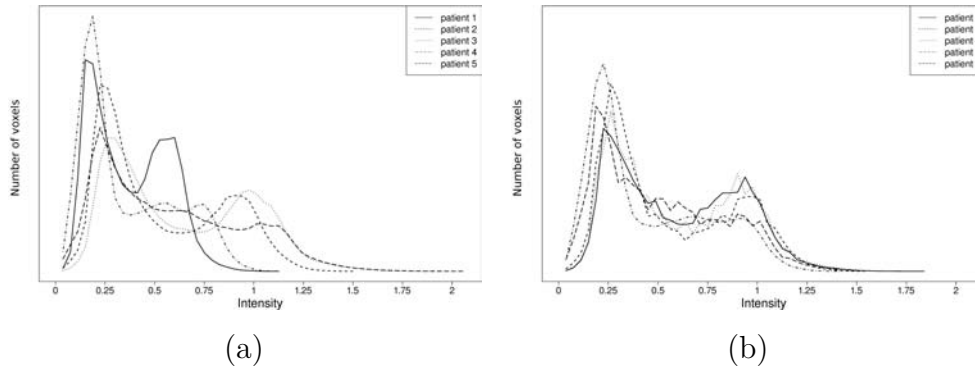


Figure 3.4: Image normalization: intensity distribution of the whole image of 5 different patients before (a) and after (b) normalization.

3.2.3 Breast-body segmentation

In this stage, the breasts are identified as the region delimited by the breast-body surface and the air-breast boundary. The breast-body surface is determined by segmenting body structures using an atlas-based voxel classification algorithm⁶⁵. The atlas-based approach uses a probabilistic atlas that provides the probability for each voxel to belong to the body (pectoral muscle, heart, lungs and thorax) and the breast as a prior information. A dedicated registration framework⁶⁵, that focuses on the body area, is used to build and map the probabilities. This registration is composed of two stages: translation transform, where the sternum is used as a landmark, and a non-rigid registration based on a B-Splines transform in a multi-resolution scheme.

The air-breast boundary is defined by a region growing algorithm applied slice by slice. A morphological dilation filter with scale 5 x 5 mm is also performed on the background segmentation to ensure that the skin between the background and the breast is removed. The probabilistic atlas-based approach is not used to compute the air-breast boundary because this approach was designed to capture anatomical variations on body area.

Figure 3.5 shows the general scheme of the breast-body segmentation framework and the results produced with its application. More details can be found in Gubern-Mérida et al.⁶⁵.

3.2.4 Breast density segmentation

In the previous breast-body segmentation step (see section 3.2.3), a general delineation of the breast is obtained. This initial breast segmentation separates the breast from the body, but it might still contain fatty tissue which is not part of the breast. In the

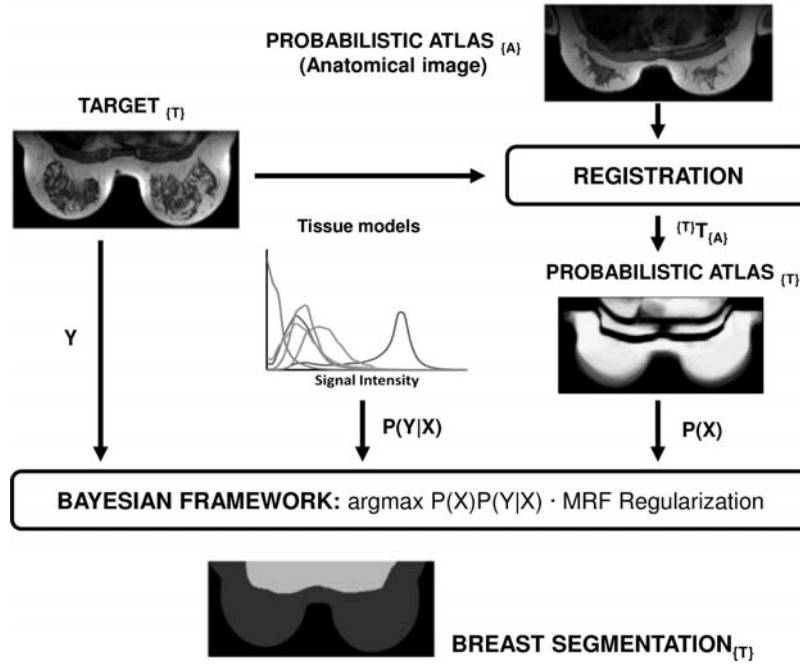


Figure 3.5: Overview of the probabilistic atlas-based breast-body segmentation approach.

breast density segmentation step, the breast volume is firstly defined for each breast independently to remove the non-breast fatty tissue that could negatively influence the estimation of breast density.

Figure 3.6 shows an illustration of the breast volume definition for the right breast. Firstly, the breasts are separated by a sagittal plane (y - z) situated in the center of the volume where the sternum was detected. Secondly, we define a coronal plane (x - y) at 2 cm posterior from the sternum landmark. Then, we exclude all the voxels which are posterior to the plane. Finally, for each breast independently, the breast boundaries are identified by detecting the superior and inferior points with maximum curvature of the central sagittal slice of the breast. Axial planes (x - z) of Fig. 3.6 define the breast boundaries of the right breast.

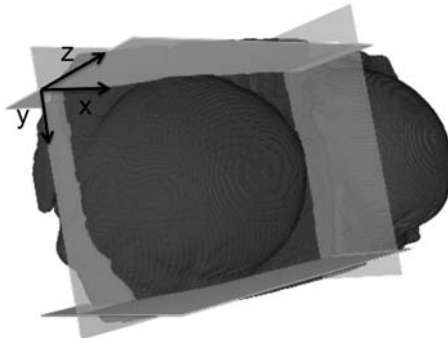


Figure 3.6: Breast volume definition for dense tissue segmentation: axial planes define the parts of the breast with highest curvature. The coronal and sagittal planes are aligned to the sternum landmark.

Once the breast volume for each breast is defined, the dense tissue is segmented using the Expectation-Maximization (EM) algorithm⁶⁶ independently on each breast. The EM is an iterative technique that maximizes the probability density function of a mixture model. In this implementation the mixture model is composed by the image intensity distributions of fatty and fibroglandular tissues modeled as Gaussian distributions, taking advantage of the good contrast between breast tissues in MRI. The signal intensity value that discriminates between fatty and fibroglandular tissue is set to the value where the likelihood ratio between the two estimated tissue models is one.

Large breasts often present folds, because the breast does not entirely fit in the coil. As can be seen in Fig. 3.7, these skin folds contain air and appear in the breast area as horizontal structures in axial slices. Voxel values of skin folds have similar image intensity values compared to dense tissue voxel values. This similarity causes skin fold voxels to be segmented as a dense tissue by the EM algorithm applied on breast volume. The erroneous segmentation of these structures can considerably bias the estimation of breast density, especially in fatty breasts. For this reason, we included a final step in our algorithm to remove such structures from the dense tissue segmentation previously generated by EM.

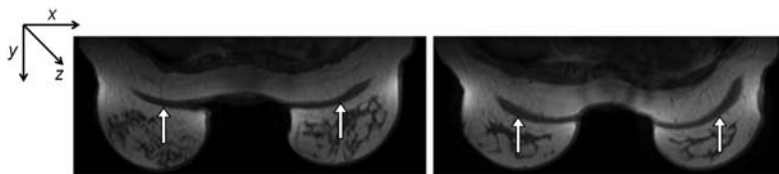


Figure 3.7: Skin fold examples.

The skin fold removal approach is applied to the whole volume that contains the dense tissue segmentation of both breasts. It is based on the observation that skin folds appear in axial slices as structures parallel to the chest wall connected to the background. Firstly, a morphological dilation operation with the scale $4 \times 4 \times 4$ mm is applied to a binary volume composed by the union of dense tissue and background segmentations. By doing so, background and skin fold voxels labeled as dense tissue are connected to each other and are detected as the largest structure in a 3D connected component analysis. Secondly, a first derivative volume $\hat{I}'_y(\mathbf{r}) = \frac{\delta}{\delta y} G_\sigma(\hat{I}(\mathbf{r}))$ is computed at Gaussian scale $\sigma = 5$ mm to enhance the structures that run parallel to the x -axis. The detection algorithm segments as skin fold the high y -gradient voxels with $\hat{I}'_y(\mathbf{r}) > \max(\hat{I}'_y(\mathbf{r}))/3$ which belong to the glandular segmentation and to the background. If skin folds are not present, only the background is found as the largest structure and no dense tissue is removed.

3.3 Experiments and results

The two main steps of the presented segmentation method were evaluated in detail using a large annotated data set. Automatic separation of body and breast was evaluated on a subset of 27 fully manually annotated MR volumes (see section 3.3.2). Fibroglandular tissue segmentation was evaluated on a set of 50 cases with manual fibroglandular tissue segmentations of 4 different experienced readers (see section 3.3.3). The complete description of the used data set is presented in the following section.

3.3.1 Material

At the Radboud University Medical Center (Nijmegen, the Netherlands), breast MRI is used for the screening of women with high familial or genetic risk. In this study we used a random subset of 50 pre-contrast coronal T1-weighted MR breast volumes from 50 different patients with no sign of malignancy within 2 years after the acquisition. The age of the screened women ranged from 23 to 76 years (45.84 ± 11.97 of average). The cases were collected from 2003 to 2009. Breast MRI examinations were performed on either a 1.5 or 3 Tesla Siemens scanner (Magnetom Vision, Magnetom Avanto and Magnetom Trio), with a dedicated breast coil (CP Breast Array, Siemens, Erlangen). Clinical imaging parameters varied; matrix size: 256 x 128 or 256 x 96; slice thickness: 1.3 mm; slice spacing: 0.625 - 1.25 mm; flip angle: 8 °, 20 ° or 25 °; repetition time: 7.5 - 9.8 ms; echo time: 1.7 - 4.76 ms. Patients were scanned in prone position. Note that, although the MR images were acquired in coronal orientation, all the annotations were performed using axial orientation and the illustrated images in this paper are axial slices.

In order to evaluate the automatic breast segmentation algorithm and construct the atlas, 27 cases were manually segmented in 7 structures by a single experienced observer. The classes are background, fatty tissue, glandular tissue, pectoral muscles, lung area and the heart. The seventh class is the “other” class and refers to the previous non-labeled voxels of the thorax. Annotations were done every 5-10 axial slices on average, and linear interpolation was applied to obtain the complete labeling. When needed, and especially for the heart, lungs and pectoral muscles, accurate manual delineation was performed within a smaller slice interval. For the manual segmentation of background and the breast, thresholding was applied on regions of interest provided by the reader. The manual segmentation of the breast dense tissue is explained further in this section. Fig. 3.8 shows an example of an MRI slice on an axial view and the manual delineation of the mentioned classes. One should note that obtaining manual annotations of the 6 different classes for the breast segmentation is a tedious and time consuming task. Each volume takes approximately 45 minutes in a dedicated breast MRI annotation

environment⁷¹. This is the reason why a subset of 27 women out of the 50 cases is used for evaluating breast segmentation.

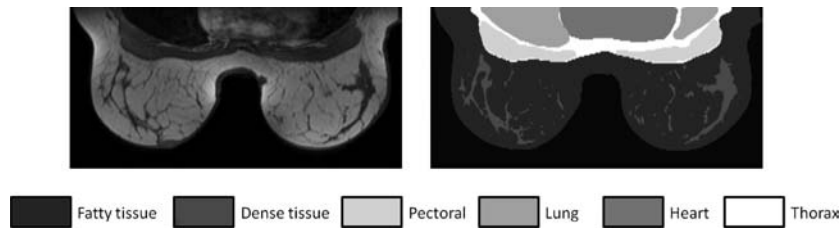


Figure 3.8: Breast MRI scan on an axial slice with manual annotations of the different structures

Independently, fibroglandular manual segmentations from different viewers were also collected for the complete data set. Firstly, an experienced observer annotated a VOI in each of the 100 breasts in the 50 cases, by coarsely encapsulating all the dense tissue while avoiding inclusion of skin folds and fatty tissue areas where signal intensity is dimmed due to the bias field. This VOI is called dense tissue mask in the paper. To delineate the dense tissue masks, the experienced observer analyzed and annotated axial slices of each MR volume using a freehand annotation tool. Secondly, 4 different readers segmented fibroglandular tissue. Each reader selected a single threshold on the dense tissue mask of each breast to discriminate between fibroglandular and fatty tissue. The thresholds were applied on bias field corrected volumes. As a result, 4 manual fibroglandular segmentations for right and left breasts were obtained for each case.

3.3.2 Breast-body segmentation

For each of the 27 fully manually segmented patients, the probabilistic atlas was built with the remaining subjects ($N = 26$) in a leave-one-out fashion. These 26 patients and their segmentations were mapped into the same reference space and the probabilities were computed⁶⁵. The same reference space was used for all the experiments. It was represented by a patient with normal appearance to minimize the differences in the registration process. This patient was not included in the evaluation set.

These 27 cases were automatically segmented and compared to their complete manual segmentations. For the evaluation of breast and body delineations we determined breast segmentation performance. Dice Similarity Coefficient (DSC)⁷², total overlap, and False Negative Fraction (FNF) and False Positive Fraction (FPF) errors are employed as done in⁴², where breast segmentation is also evaluated. DSC is widely used in the literature to evaluate segmentation agreement and is also referred to as mean overlap. The total overlap is computed as the sensitivity. The FNF error gives the fraction of the breast reference segmentation that is not segmented as the breast by the automatic segmentation. Similarly, the FPF error gives the fraction of the automatic

segmentation of the breast that is not segmented as the breast by the reference segmentation. More details about these measures can be found in Klein et al.⁷³. Furthermore, since similarity between voxel values of pectoral muscle and fibroglandular tissue is the reason why breast-body separation is not trivial, we provide a complementary evaluation: the fraction of fibroglandular tissue voxels of the reference segmentation labeled as body in the automatic segmentation of the breast. We refer to this evaluation measure as Dense Tissue Error (DTE). To compute this metric, the fibroglandular tissue reference segmentation was built by combining the 4 manual dense tissue segmentations of each observer using STAPLE⁷⁴. The STAPLE algorithm fuses a collection of manual segmentations maximizing the accuracy assessment of each expert. The DTE metric was not previously used in the literature, but we consider it relevant to quantify the segmentation error in combination with other overlap measures.

Figure 3.9 shows a box plot with the distribution of overlap and error metrics used for breast segmentation evaluation. The average DSC value was 0.94 ± 0.03 (mean \pm sd), the average total overlap was 0.96 ± 0.02 (mean \pm sd), the average FNF was 0.04 ± 0.02 (mean \pm sd), the average FPF was 0.07 ± 0.06 (mean \pm sd) and the average DTE was 0.02 ± 0.03 (mean \pm sd). The DTE was useful to observe the influence of pectoral muscle in the breast-body surface segmentation. The majority of the fibroglandular voxels automatically misclassified as body (93.75%) belonged to pectoral muscle, which is included in the body. Accurate delineations are achieved as is illustrated in Fig. 3.10.

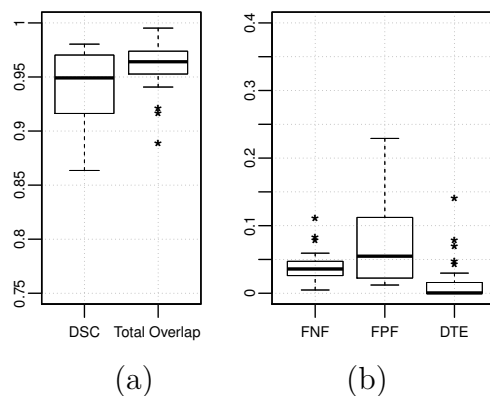


Figure 3.9: Breast segmentation performance: (a) distribution of overlap DSC and total overlap metrics and (b) distribution of error overlap False Negative Fraction (FNF), False Positive Fraction (FPF) and Dense Tissue Error (DTE).

The sternum localization and its influence on the breast segmentation stage was also evaluated. Each case was annotated with a manually centered sternum point. Manual landmarks of the outer edge of the sternum were compared to the sternum localizations determined automatically. An average error of 1.51 ± 1.34 mm (1.16 ± 1.03 voxels) was obtained, which shows the validity of our sternum detection approach. To evaluate the impact on the final segmentation, the automatic breast-body segmentation approach

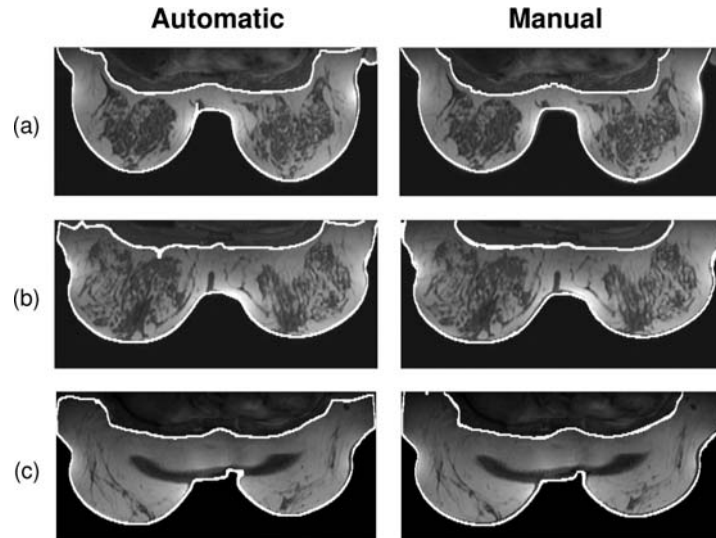


Figure 3.10: Intermediate slices from 3 different cases and the breast contour obtained from the segmentations.

was also tested without incorporating the sternum landmark detection. When the sternum point was incorporated, an improvement was observed in the performance of pectoral muscle segmentation. A DSC mean value of 0.75 ± 0.09 (mean \pm sd) was obtained for pectoral muscle segmentation. When the sternum point was not incorporated, the DSC value decreased to 0.71 ± 0.12 (mean \pm sd). In fatty cases, where the registration is more complicated because of the large amount of fatty tissue between the body and the background (ie. case (b) of Fig. 3.10), the localization of the sternum appears to be essential as can be seen in the example in Fig. 3.11.

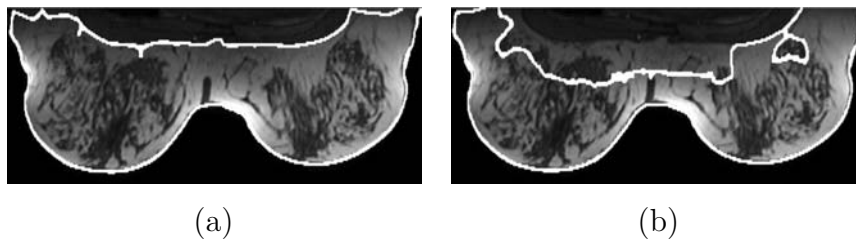


Figure 3.11: Breast segmentation of the case 7 with (a) and without (b) the use of sternum landmark in registration.

3.3.3 Breast density segmentation

In order to evaluate the breast density segmentation, several experiments were performed. Firstly, inter-observer variability was studied analyzing the 4 manual segmentations of the 50 cases (100 breasts). Secondly, to show the validity of the EM algorithm for fibroglandular tissue segmentation, we also segmented dense tissue of the 50 cases using the EM algorithm on the manually segmented dense tissue masks employed in

manual thresholding. Finally, breast density segmentations from the complete data set were obtained using the presented fully automatic approach, which includes all steps: automatic breast segmentation and density segmentation using EM.

STAPLE⁷⁴ was used to fuse manual segmentations and build multiple fibroglandular tissue reference segmentations. To evaluate inter-observer variability, a leave-one-out strategy was followed and each manual segmentation was compared to a reference segmentation created by fusing manual delineations of the 3 remaining readers. In total, four fibroglandular tissue reference segmentations were created for each case. To evaluate our approach, automatic segmentations were compared to each of the four reference segmentations and the average was taken. DSC, total overlap, FNF and FPF were chosen to quantify the performance of breast density segmentation. Two-sided paired Wilcoxon test at 95% confidence level was used to determine statistical significance.

The results presented in Table 3.1 and in Fig.3.12 indicate that our EM algorithm is comparable to manual thresholding when applied on a manually defined dense tissue area. In terms of DSC, which is the metric used by Wu et al.^{63,64} to evaluate the performance of their fully automatic fibroglandular tissue segmentation algorithms, it outperforms segmentations of viewer 3 (p-values < 0.025) and differences are not significant when compared to viewer 1, 2 and 4 (p-values of 0.20 and 0.54 and 0.54 respectively). When the fully automatic method is applied, the performance decreases to 0.80. However, automatic segmentations have similar performance compared to that obtained by viewer 3 (p-value of 0.65). Figure 3.13 shows reference and automated fibroglandular segmentations on two axial slices from the best and worst automatic segmented cases. In the worst case example, the pectoral muscle is segmented as fibroglandular tissue due to registration errors and the skin fold was not totally removed. Although the skin fold is still present in this example, overall, the skin fold removal step appeared to be important: The results obtained by the proposed fully automatic approach, which includes skin fold removal, significantly outperforms the fibroglandular tissue segmentations obtained without detecting such structures (average DSC values of 0.80 ± 0.12 (mean \pm sd) and 0.75 ± 0.14 (mean \pm sd) respectively, with p-value < 0.025).

Finally, the volume of dense tissue was computed using the fibroglandular tissue segmentations of the automated approach and the reference segmentations. For each case and segmentation, dense volume estimation of the right and left breast was compared. Figure 3.14 illustrates the correlation (0.93 and 0.91) between right and left breast densities for (a) fully automatic and (b) reference segmentations. Numbers indicate the case identifier. Assuming that the right and left breasts tend to have the same density, the similarity of volumetric estimates obtained for left and right breasts demonstrates the consistency of the method we developed. Similarly, correlation between breast den-

Table 3.1: Average DSC, average total overlap, average FNF and average FPF (mean \pm sd) obtained from manual segmentations of the 4 different viewers, from Expectation-Maximization algorithm applied on manually defined dense tissue area (EM mask) and, from EM algorithm applied on automatic defined breast volume (Automatic).

	DSC	T. Overlap	FNF	FPF
V1	0.84 ± 0.11	0.77 ± 0.17	0.23 ± 0.17	0.04 ± 0.07
V2	0.88 ± 0.08	0.93 ± 0.11	0.07 ± 0.11	0.15 ± 0.12
V3	0.80 ± 0.15	0.95 ± 0.08	0.05 ± 0.08	0.28 ± 0.21
V4	0.85 ± 0.12	0.77 ± 0.17	0.23 ± 0.17	0.02 ± 0.03
EM mask	0.86 ± 0.10	0.89 ± 0.11	0.11 ± 0.11	0.14 ± 0.15
Automatic	0.80 ± 0.13	0.85 ± 0.11	0.15 ± 0.11	0.22 ± 0.18

sity measures computed from automatic and reference segmentations was substantial, reaching breast-based and case-based correlation ratios of 0.89 and 0.92, respectively (see Fig.14(c)).

3.4 Discussion and conclusions

The segmentation framework we developed to automatically compute breast density by segmenting breast and fibroglandular tissue in breast MRI shows good performance for both automatic segmentation tasks. The presented method is a combination of image processing techniques. Intra- and inter-patient signal intensity variability is initially corrected. Secondly the breast is segmented by automatically detecting body-breast and air-breast surfaces using an atlas-based approach. Subsequently, the fibroglandular tissue is segmented by EM algorithm and skin-fold removal.

Different stages of our method have been validated. Automatic segmentations of breast and fibroglandular tissue were compared to manual segmentations. Overall, qualitative and quantitative results show that the obtained segmentations are of high quality in most cases. With respect to the fibroglandular tissue segmentation, good correlation between breast density estimates obtained in the right and left breast indicates consistency and robustness of our approach.

We observed that the inclusion of the sternum landmark in the registration framework improves the segmentation of the breast-body surface. This is important because erroneous inclusion of pectoral muscle tissue in the breast is a major source of errors in methods aimed at automated segmentation of fibroglandular tissue. The sternum landmark was detected automatically and this detection was reliable since results we obtained were similar to manual annotations.

Compared to previous work found in the literature, for breast segmentation, our

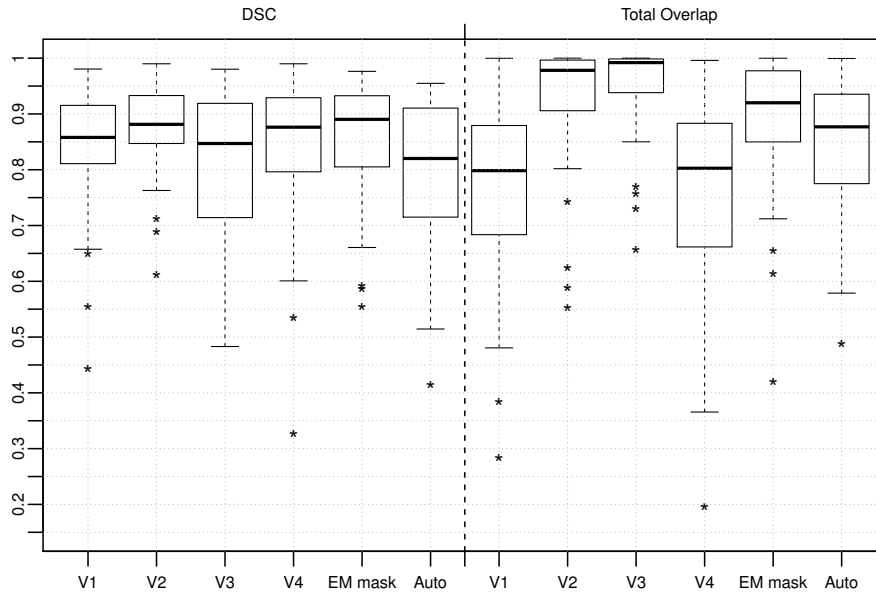


Figure 3.12: Fibroglandular tissue segmentation: distribution of DSC and total overlap measures computed from manual segmentations of the 4 different viewers (V#), from Expectation-Maximization algorithm applied on manually defined dense tissue area (EM mask) and, from EM algorithm applied on automatic defined breast volume (Automatic).

performance metric values are higher than the reported by Gallego-Ortiz et al.⁴². With our approach we achieved an average DSC value of 0.94 ± 0.04 (mean \pm sd) and an average total overlap of 0.96 ± 0.02 (mean \pm sd). Their breast segmentation framework obtained an average DSC value of 0.88 ± 0.05 (mean \pm sd) and an average total overlap of 0.89 ± 0.05 (mean \pm sd). However, Gallego-Ortiz et al. validated their approach on 409 MRI cases, which is larger than the 50 used in our work, and reported computational time of less than a minute per volume. Our approach takes approximately 8 minutes computed on a 4 Intel(R) Core(TM) i5-2400 CPU 3.10GHz processors system with 12 GB RAM. Most of this time is spent by our algorithm to register and map the probabilistic atlas to the target image (≈ 6 min.). However, one should note that the current implementation of this step is single-threaded. Therefore, the computational time can be substantially decreased by multithreading the task.

For the automatic fibroglandular tissue segmentation, we obtained an average DSC value of 0.80 ± 0.13 (mean \pm sd). The DSC value is higher than the values of 0.73 and 0.67 reported by Wu et al.^{63,64}. These authors also reported a value of 0.93 for case-based fibroglandular tissue volume correlation between automatic and manual segmentation⁶⁴. In our study, we obtained a case-based fibroglandular tissue volume correlation of 0.92.

It should be noted that not too much importance should be given to these comparisons, since the segmentation results were determined on different data sets. The composition of these sets may have a strong influence on the overall results, especially

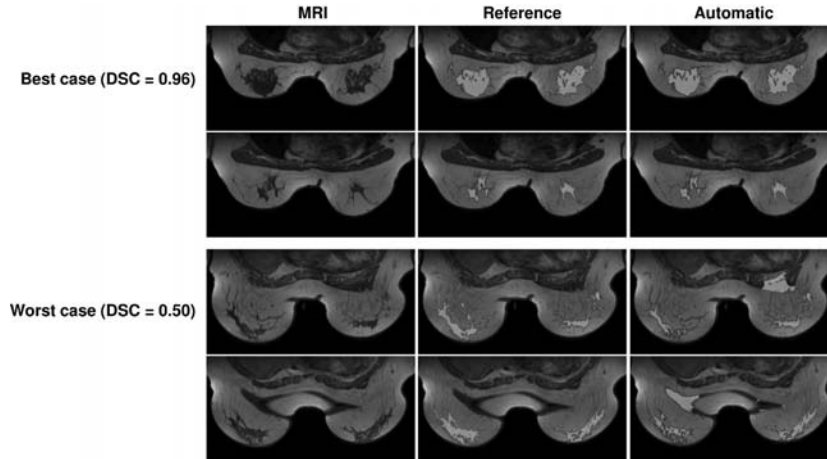


Figure 3.13: Fibroglandular tissue segmentation results: comparison of reference and automated segmentation of two axial slices from the best and worst automatic segmented cases.

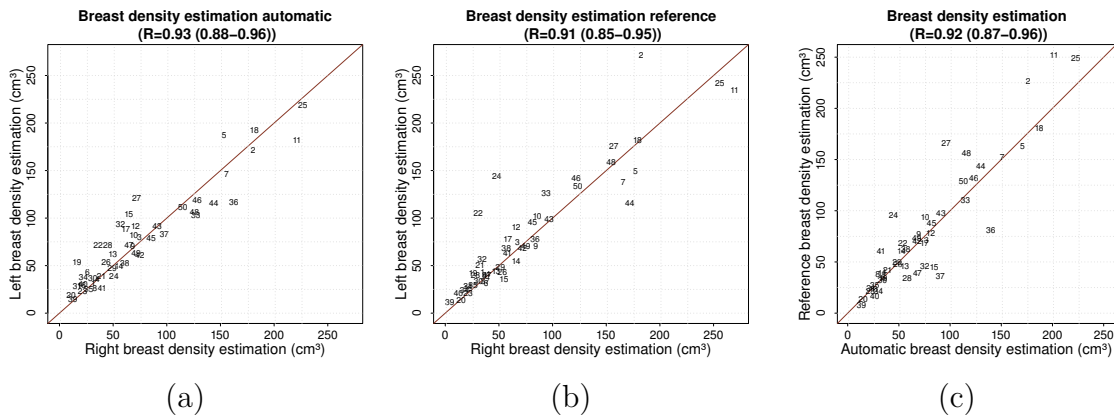


Figure 3.14: Correlation between right and left breast density estimations using (a) fully automatic and (b) reference segmentations and (c) case-based correlation between breast density estimations from automatic and reference segmentations. Numbers refer to the identifier of each case.

when sets are small. Data sets represent different populations and the way manual annotations are made may also cause differences. For instance, we observed that the area covered by the manual breast segmentations used in our experiments is larger than the area covered by the manual segmentations used by Gallego-Ortiz et al.⁴². These differences are caused by the difficulty of determining the anatomical extent of the breast, in particular the lateral zone where the breast is connected to the chest wall.

Using the presented method, we found that most of the remaining errors in automated fibroglandular tissue segmentation are due to an inaccurate definition of the breast-body surface. In the experiments, in three of the cases the breast segmentation algorithm classified more than 6 % of dense tissue voxels as body ($DTE > 0.06$), which indicates that a substantial amount of body voxels were erroneously labeled as dense tissue. This error was observed in cases where the registration was not able to compensate for differences between the target and the atlas. Hence, probability distributions

of body and breast were not correctly mapped leading to errors in the final segmentation. To improve results in such cases, the use of a multi-probabilistic atlas approach can be considered, in which different probabilistic atlases would be created for different populations of breast MRI cases, grouped based on shape and/or patient age. For a given case, the most similar atlas to the target image would be chosen to proceed with the segmentation. A disadvantage of this approach, however, is that it may lead to a strong increase in computational load.

The use of a different threshold to segment fibroglandular tissue for each breast appeared to be important. During the development of our algorithm, we found that the use of a single threshold for both breasts produced a strong asymmetry between left and right breast density estimations. This asymmetry was caused by differences in signal intensities between left and right breasts. We considerably reduced it and improved overall performance of the presented method by applying the N3 bias field correction method and by segmenting each breast independently. We also observed that the bias field is not completely removed after correction in cases with large breasts. The bias field causes lower signal intensity values in fatty tissue areas of the base breast and this fatty tissue is incorrectly segmented as fibroglandular tissue. To overcome this limitation the effect of applying other bias field correction algorithms⁷⁵ will be studied in the future.

The method in our study has been developed and evaluated using data collected in only one clinical center. Therefore, robustness against differences in population and image acquisition systems could not be investigated. However, we are confident that, with minor adjustments, the proposed segmentation framework for breast MRI is suitable for application to data sets obtained in different institutions and under different conditions, since our dataset contained images obtained on different scanners with different clinical imaging protocols and our method only uses a few parameters. These parameters are related to sternum detection and skin-fold removal. Experimentally we found that the performance of our method did not critically depend on the values of these parameters.

The probabilistic atlas we used captures anatomical information. Since it is learned from a selected population, care must be taken that the atlas is representative for the population the method is applied to. The atlas is not dependent on the image acquisition protocols, but obviously it needs to be rotated and scaled to the correct coordinate frame. It is remarked that the use of different breast coils may lead to some variations that are not represented in an atlas constructed from a single acquisition system. However, since this predominantly affects the atlas values of breast tissue, this does not affect our method, in which the atlas is only used for the breast-body surface segmentation.

To summarize, in this paper we have presented a new fully automatic approach to measure the volume of dense tissue in breast MRI. The previously published segmentation methods either required manual intervention, were evaluated on small data sets, or were focused only on breast-body separation. Evaluation on a substantial data set of fifty cases with multiple manual annotations has shown that our approach yields fairly accurate automatic segmentations. Therefore, we believe that our segmentation method can serve as a reliable tool in breast density studies, such as investigating the correlation between breast density measurements obtained from MRI and mammograms. Although there is no consensus on the role of breast density as an individual risk factor for breast cancer, population based studies have shown that there is a strong relation between breast density and breast cancer occurrence. The presented segmentation framework, as well as its different stages independently, can be also applied in other clinical applications for breast cancer. For instance, fibroglandular tissue morphology and distribution patterns can be analyzed to investigate their relationship with risk of developing cancer, and automated segmentations may be used in computer aided diagnosis programs.

Validation of volumetric breast density estimations from FFDM using breast MRI

4

Albert Gubern-Mérida, Michiel Kallenberg, Bram Platel, Ritse M. Mann, Robert Martí, Nico Karssemeijer

Original title: Volumetric breast density estimation from Full-Field Digital mammograms: a validation study

Published in: PLoS ONE (2014); 9(1):e85952

Abstract

Objectives: To objectively evaluate automatic volumetric breast density assessment in Full-Field Digital Mammograms (FFDM) using measurements obtained from breast Magnetic Resonance Imaging (MRI).

Material and methods: A commercially available method for volumetric breast density estimation on FFDM is evaluated by comparing volume estimates obtained from 186 FFDM exams including mediolateral oblique (MLO) and cranial-caudal (CC) views to objective reference standard measurements obtained from MRI.

Results: Volumetric measurements obtained from FFDM show high correlation with MRI data. Pearson's correlation coefficients of 0.93, 0.97 and 0.85 were obtained for volumetric breast density, breast volume and fibroglandular tissue volume, respectively.

Conclusions: Accurate volumetric breast density assessment is feasible in Full-Field Digital Mammograms and has potential to be used in objective breast cancer risk models and personalized screening.

4.1 Introduction

Breast density has been identified as an important risk factor for developing breast cancer. Studies have reported that the risk of getting breast cancer in women with high breast density is four to six times as large as in women with low breast density^{10,58,76}. Additionally, sensitivity of mammography screening is severely impaired in women with high density, since the presence of heterogeneous or extreme dense tissue patterns may obscure suspicious lesions. For this reason, the risk of missing cancers in screening programs increases with density^{5,7,77}. Personalization of screening protocols, involving adjunct imaging modalities for women who are currently not adequately screened, has been suggested to circumvent this problem. Such protocols should include risk assessment based on models including family history and breast density biomarkers¹¹.

To develop such models, it is important to objectively measure breast density. Most studies to date have been performed using subjective visual measurements based on the 4-class Breast Imaging Reporting and Data Systems (BI-RADS)⁷⁸, which is used in current clinical practice, or on a visual thresholding technique using dedicated software, such as Cumulus⁷⁹. Both are essentially 2D measurements that determine the area of dense tissue projected in mammograms. Fully automatic methods for area based breast density measurements have been proposed to take subjectivity away⁸⁰⁻⁸³. However, area based measurements do not take the thickness of dense tissue into account. This is a limitation since it is biologically more plausible that breast cancer risk is related to the volume of dense tissue in the breast rather than to its projection^{58,84,85}.

To overcome this limitation, methods for volumetric breast density estimation from mammograms have been proposed^{45,46,86-89}. These methods are based on a physics based model of the X-ray image acquisition process and assume that the breast tissue consists of two types of tissue: fat and parenchyma. By knowing the X-ray attenuation of these tissues, tissue composition at a given pixel can be computed. Initially, researchers have struggled to successfully apply this approach to digitized film mammograms. However, with the introduction of Full-Field Digital Mammograms (FFDM), the development of robust methods and commercial products became possible. Those can be applied to raw (unprocessed) FFDM data, which is made available by all modality manufacturers. Unfortunately, though, raw data is often not archived in clinical practice.

The performance of volumetric breast density estimation methods has been evaluated in several studies. To determine robustness and consistency, comparisons have been made of breast density estimates in the left and right breasts, and in mediolateral oblique (MLO) and cranial-caudal (CC) exposures of the same breast^{45,86}. One would expect to find similar values in CC and MLO views and, in regular cases without abnor-

malities, breast density in the left and right breast should be highly correlated. Other studies compared volumetric estimates to BI-RADS density scoring^{90,91}. These previously mentioned validation strategies may not reveal systematic errors, while subjective BI-RADS scorings are coarse and inaccurate by nature and are only useful to determine large errors of the automated methods. Comparison of breast density estimates from FFDM to reference standard measurements obtained from three-dimensional imaging modalities, such as Magnetic Resonance Imaging (MRI) and Computed Tomography (CT), is arguably the most objective and complete validation method^{45,46,92,93}. The volume of dense breast tissue can accurately be derived from MR and CT images, as these are 3D acquisitions and no projection is involved. However, quantification of the volume of dense breast tissue is a time consuming task when done by means of manual segmentations because it requires segmentation of 3-dimensional data. For this reason we use computer algorithms to obtain breast density measurements.

In this paper, we evaluate a method for measuring volumetric breast tissue estimates from digital mammograms^{45,46}. We specifically studied the performance of the method for determination of fibroglandular tissue volume, breast volume, and volumetric breast density by comparing its results to volume estimates that were obtained from breast MRI data.

4.2 Materials and methods

4.2.1 Dataset

Ethics Statement: According to the Dutch Medical Research Involving Human Subjects Acts (WMO), retrospective studies using only patient records do not require a formal medical ethics review and informed consent is not needed. The need for signed informed consent was waived by the Independent Review Board (IRB). This was confirmed with the local medical ethical committee and can be read at www.ccmo-online.nl. The presented study complies with the Dutch Data Protection Authority requirements on the use of patient data.

In the Radboud University Medical Center (Nijmegen, the Netherlands), breast MRI and mammography are used for screening of women with high familial or genetic risk. We included studies for which breast MRI data and FFDM were available with time interval between these exams of less than two months. We obtained 250 MRI volumes and 928 MLO and CC images from FFDM exams from 250 studies (132 different women). Mean time between MRI and FFDM acquisitions was six days. CC views were not available in some cases. All exams were performed between December 2000 and December 2011. The age of the screened women ranged from 24 to 77 years, and was 46.5 ± 11.10 years on average.

The digital mammograms used in the study were acquired on a GE Senographe 2000D or on a GE Senographe DS using standard clinical settings, including the use of an anti-scatter grid. Breast MRI examinations were performed on 1.5 or 3 Tesla scanners (Magnetom Vision, Magnetom Avanto and Magnetom Trio, Siemens) with a dedicated breast coil (CP Breast Array, Siemens). In this study we used pre-contrast T1-weighted MR volumes.

4.2.2 Breast density quantification

In this study, volumetric breast density, breast volume and fibroglandular volume estimates were obtained from FFDM and MRI data. Volumetric breast density refers to the percentage of breast density, computed by dividing the fibroglandular tissue volume by breast volume.

Volumetric estimates from 250 FFDM studies were obtained using Volpara 1.4.3 (Mātakina, Wellington, New Zealand), which is FDA-approved fully automated software to estimate volumetric breast density. The Volpara method is an extension of the algorithm presented by van Engeland et al.⁴⁵. In particular, it incorporates a more detailed physics model including scatter components as described by Highnam et al.⁸⁷, and it uses a more advanced method to determine a reference region of fatty tissue. This reference region is used for calibration, and allows computation of fibroglandular tissue thickness at every pixel in the image. Breast volume is determined using a geometric model in which the periphery of the compressed breast is modeled by semi-circular cross sections, using the breast thickness measurement provided by the acquisition system in the image header.

Volumetric measurements from MRI were obtained using a multi-probabilistic atlas-based segmentation method based on the work of Gubern-Mérida et al.^{50,65}. In short, the breast MRI segmentation method initially corrects the bias field and normalizes signal intensities among patients. Secondly, probabilistic atlases, which capture the anatomic variation of the pectoral muscle and chest wall, are used to segment the breast. A probabilistic atlas is a volume that contains the complete spatial distribution of probabilities of voxels to belong to one or more organs⁶⁵. Finally, the fibroglandular tissue is segmented in each breast independently using automatic thresholding. In this work, this method was used to automatically segment breast and fibroglandular tissue in the 250 MRI studies. A radiologist with expertise in breast imaging carefully reviewed all slices of the segmentations and approved 186 (74.4%) MRI studies with segmentations to be suitable for the use as a reference standard for validation of FFDM density measurements. The other 64 (25.6%) studies were excluded from the study. The field of view of 5 of the excluded cases did not entirely cover the breast. In the rest of the excluded cases we observed that the main reason for the MRI segmentation

failure was the presence of artifacts or bias field remaining after correction. These signal intensity distortions negatively affected the segmentation process.

4.2.3 Validation

The validation process is represented in Fig. 4.1. The Volpara method was validated on 186 FFDM exams including 680 mammographic views. The Pearson's correlation coefficients between volumetric measures obtained from FFDM and volumetric measures obtained from MRI were calculated per breast and per study. The volumetric estimations per breast from FFDM were averaged over available measures of CC and MLO views for each breast independently. Measures per study were computed by averaging right and left breast volumetric estimates. Because of the log-normal distribution of the data, correlation coefficients were computed after converting the measurements using the natural logarithmic transform⁹¹.

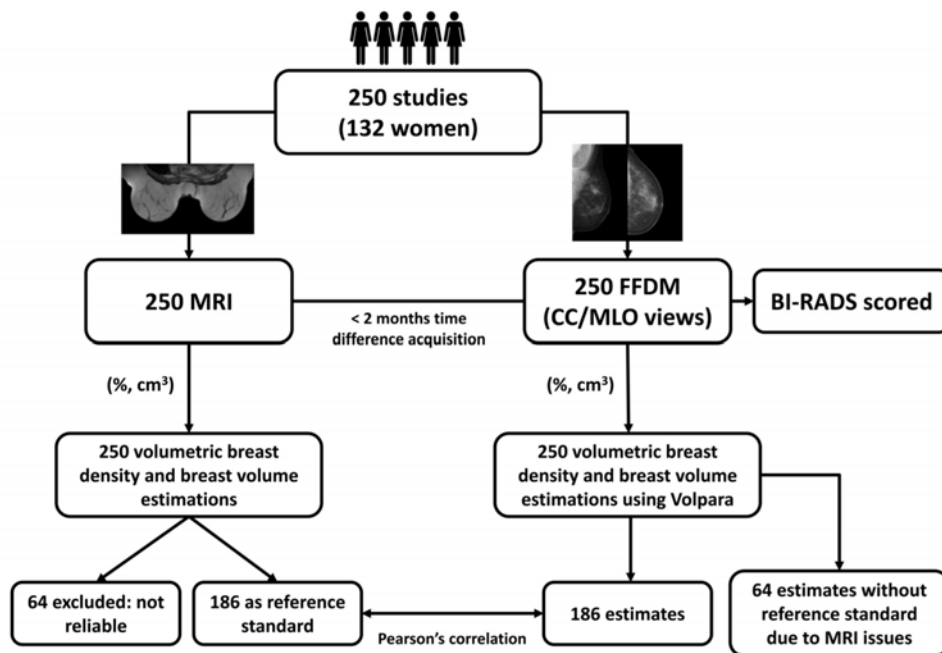


Figure 4.1: Schematic overview of the validation process.

Scatter plots are used to visualize the comparison between breast volumetric estimations. Volpara Density Grade (VDG) thresholds are also shown for volumetric breast density estimates obtained from FFDM. The VDG is a grading system that maps the percent density output of Volpara into four categories similar to the BI-RADS density score. The ranges of the percentage of dense tissue for VDG 1, 2, 3 and 4 are 0 – 4.5%, 4.5 – 7.5%, 7.5 – 15.5% and 15.5% and up, respectively⁹⁴.

BI-RADS density scoring (1 to 4) was also performed on the 250 FFDM studies. Each study was classified as (1) fatty, (2) scattered dense, (3) heterogeneously

dense or (4) extreme dense by a breast radiologist. Volumetric breast density measurements obtained from FFDMs and MRI, computed per study, were compared to its BI-RADS category provided by the radiologist and the Spearman Ranked correlations were computed for each modality. Finally, to quantify the concordance between VDG and BI-RADS density score, the weighted kappa with quadratic weights coefficient was measured.

4.3 Results

Table 4.1 summarizes the results obtained in this validation study. Figure 4.2 shows the relation between percentage of volumetric breast density from mammograms and MRI data per breast (a) and per study (b). Correlations per breast and per study are 0.91 and 0.93, respectively. Figure 4.3 shows the relation between breast volume estimates from mammograms and MRI data. Per breast (a) and per study (b) correlations are both 0.97. Additionally, Fig. 4.4 shows the relation between fibroglandular tissue volume estimates from mammograms and MRI data. Correlation per breast (a) is 0.84 and correlation per study (b) is 0.85.

Table 4.1: Summary of the dataset and the results obtained in this study. IQR=inter-quartile range, *=Pearson correlation coefficient, +=Spearman Ranked correlation coefficient, -=weighted kappa with quadratic weights coefficient.

Number of studies	186	
Number of mammographic views	680	
Number of breasts	353	
	FFDM (Median (IQR))	MRI (Median (IQR))
Volumetric breast density (%)	11.90 (12.86)	13.55 (17.15)
Breast volume (cm ³)	551.95 (405.32)	643.16 (439.56)
Fibroglandular tissue volume (cm ³)	60.45 (50.36)	76.27 (72.20)
	Per breast	Per study
Volumetric breast density correlation		
- FFDM - MRI	0.91*	0.93*
- FFDM - BI-RADS	-	0.78+
- MRI - BI-RADS	-	0.77+
- VDG - BI-RADS	-	0.40-
Breast volume correlation		
- FFDM - MRI	0.97*	0.97*
Fibroglandular tissue volume correlation		
- FFDM - MRI	0.84*	0.85*

Overall, high correlation between FFDM and MRI measurements is observed. However, results indicate that Volpara tends to underestimate breast density in dense breasts

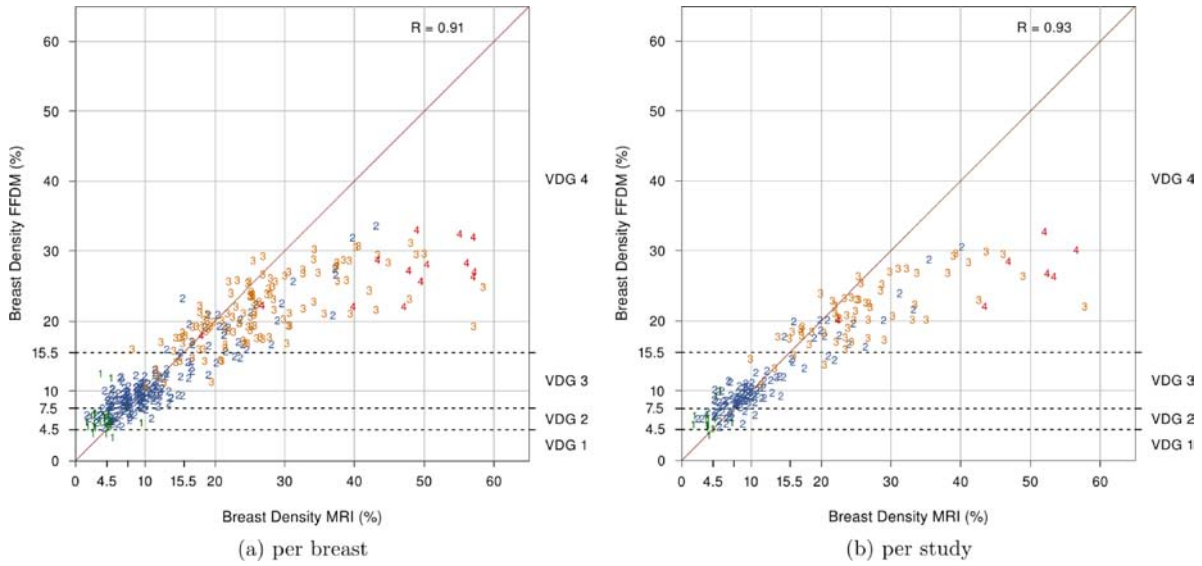


Figure 4.2: Comparison of percentage of breast density from MRI and FFDMs (a) per breast ($n = 353$) and (b) per study ($n = 186$). Each point is labeled with the BI-RADS score. VDG 1, 2, 3 and 4 refer to Volpara Density Grade breast density percentage ranges.

compared to MRI. Correlation drops for volumetric breast density measurements classified within the VDG 4 range.

Furthermore, Fig. 4.5 shows the association between volumetric breast density estimates and BI-RADS category. The estimates are obtained from FFDMs on Fig. 4.5(a), and obtained from MRI on Fig. 4.5(b). Spearman Rank correlation coefficients are 0.79 and 0.78 for FFDM and MRI, respectively. The reported correlations are not statistically significantly different (p -value = 0.71, two-tailed z -test). Following the trend observed before, volumetric breast density estimates are larger when obtained from MRI than when computed on FFDMs. The median estimates obtained with Volpara range from 5.66%, in the lowest BI-RADS category, to 26.69%, in the top category. Median estimates obtained from MRI data range from 3.80% to 52.00%. Figure 4.6 shows the number of studies scored with BI-RADS categories 1, 2, 3 and 4 for (a) the initial dataset and for (b) the dataset after excluding studies with poor MR segmentations. Finally, Table 4.2 shows the confusion matrix for the VDG using the Volpara method versus BI-RADS density score given by the breast radiologist. The weighted kappa with quadratic weights statistic was 0.40.

4.4 Discussion

In this study we have presented a validation of Volpara 1.4.3 (Mātakina, Wellington, New Zealand), which is a commercially available method for assessing volumetric breast density on FFDM. Volpara has been evaluated on 186 studies including 680 mammographic views of 353 breasts in total. Volumetric estimates obtained from FFDM have

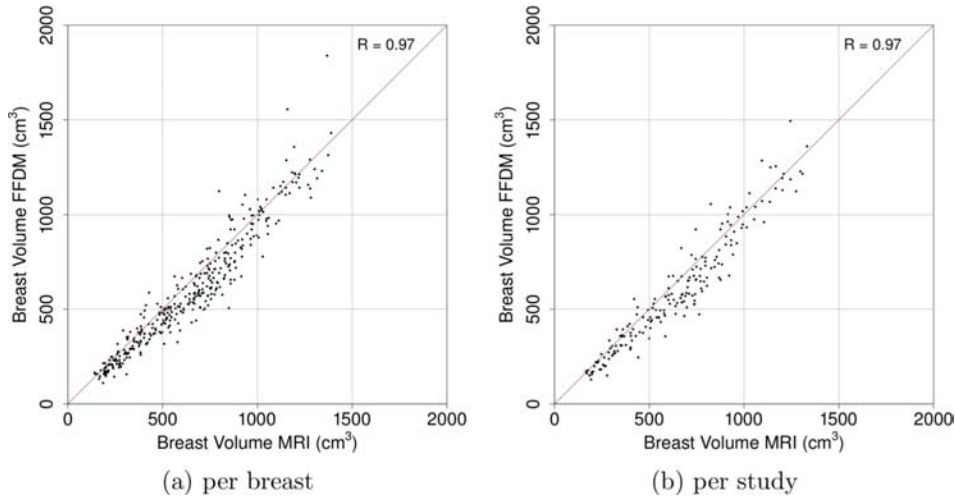


Figure 4.3: Comparison of breast volume obtained from MRI and FFDMs per (a) breast ($n = 353$) and (b) per study ($n = 186$).

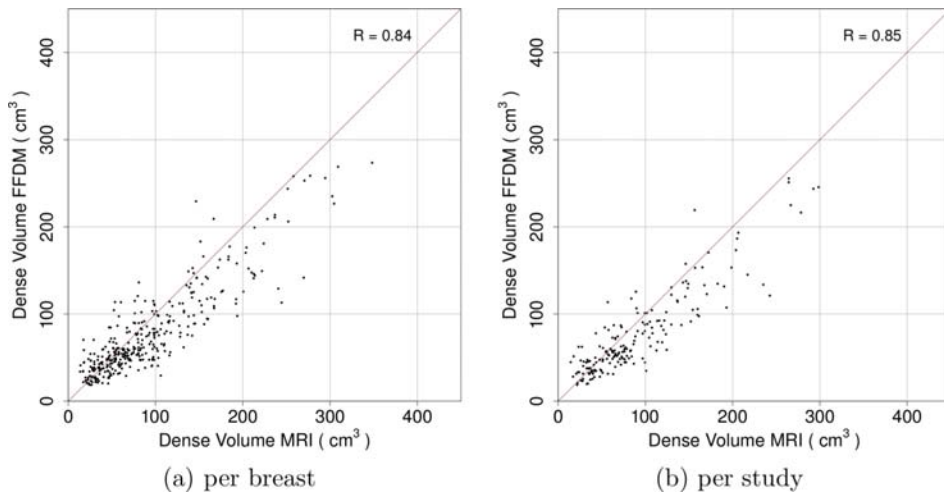


Figure 4.4: Comparison of fibroglandular tissue volume obtained from MRI and FFDMs (a) per breast ($n = 353$) and (b) per study ($n = 186$).

been compared to objective reference standard measures computed from MRI. Volumetric breast density and breast tissue volume values obtained with Volpara present high correlation when compared to MRI measurements. To date, this is the largest validation study that compares volumetric breast density estimates from FFDM to reference standard measurements obtained from MRI, a 3D imaging modality.

In previous work, Wang et al.⁹³ used a dataset of 123 patients and also compared volumetric measurements obtained from FFDM to estimates obtained from MRI. Correlations for breast volume, fibroglandular tissue volume and volumetric breast density were 0.94, 0.62 and 0.71, respectively. We found higher correlation values than the ones reported in their work ($R=0.97$, $R=0.85$ and $R=0.93$ for breast volume, fibroglandular tissue volume and volumetric breast density, respectively). Van Engeland et al.⁴⁵ also compared density estimates from FFDM to estimates from MRI in a small study includ-

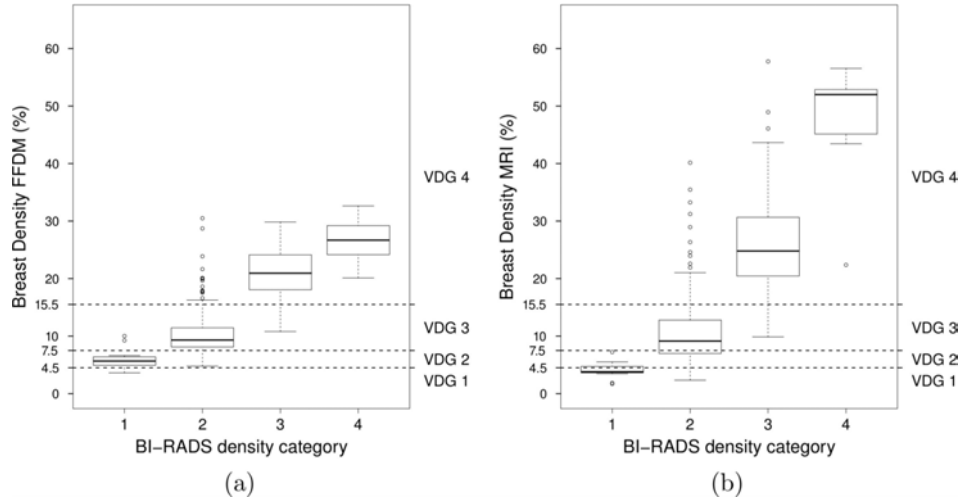


Figure 4.5: Association between volumetric breast density estimates per study and BI-RADS category.

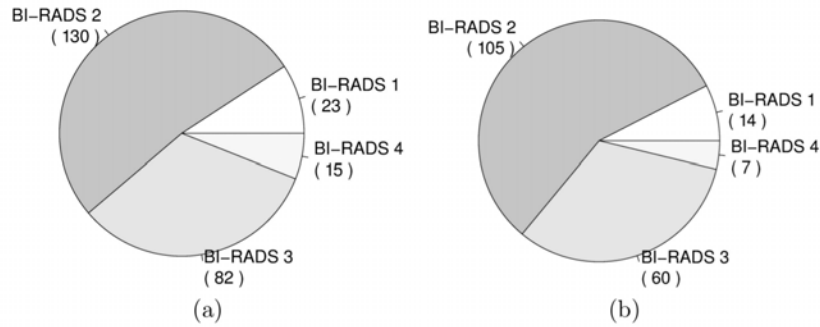


Figure 4.6: Frequency of studies scored with BI-RADS categories 1, 2, 3 and 4 for (a) the complete dataset ($n = 250$) and (b) for the cases of the dataset with reference standard estimates ($n = 186$).

ing 22 patients, but only reported correlation between fibroglandular tissue volume from mammograms and from MRI data. The correlation was 0.97. In our study we found a lower correlation between fibroglandular tissue volume from FFDM and from MRI ($R=0.84$). In previous studies, Volpara was also compared to semi-automatic area-based density measurements. High correlation between the volumetric breast density obtained with Volpara and area-based percentage density using Cumulus was found ($R=0.85$)⁹¹. Care should be taken when comparing the correlation coefficients obtained in this work to the values reported in similar studies; these similar studies were performed on different datasets. In our study, the dataset was mostly composed of pre-menopausal women participating in a high-risk screening program. In this dataset, a different distribution of breast density may be expected when compared to breast density distributions of other datasets, since there are many factors that influence breast density (such as age and use of hormone replacement therapy). On the other hand, we may assume that the appearance of fibroglandular tissue itself in our study group is similar to that in other studies, since there is no evidence that breast density patterns in women in a high risk

Table 4.2: Volpara Density Grade (VDG) versus BI-RADS density score from a breast radiologist.

		BI-RADS				
		1	2	3	4	
Volpara (VDG)	1	2	0	0	0	2
	2	10	19	0	0	29
	3	2	70	5	0	77
	4	0	16	55	7	78
		14	105	60	7	186

population differ from those in the general population.

Compared to volumetric measurements obtained from MRI, results show that Volpara tends to underestimate breast density in very dense breasts. This effect has been also observed in other methods for volumetric breast density estimation^{45,95}. Like Volpara, these methods are also based on a physics-based image model and, to predict fibroglandular tissue thickness, use a set of pixels of the breast that belong to fatty tissue as an internal reference. The selection of the internal reference is more complex in dense breasts than in fatty breasts, which affects the calibration of fatty tissue attenuation and leads to breast density underestimation. However, the breast density underestimation in dense cases does not seem to affect the final VDG categorization. We observed that the cases with the largest negative difference between estimates from FFDM and MRI obtained a volumetric breast density estimate from FFDM greater than 15% and were classified as VDG 4.

Compared to BI-RADS density scores given by a breast radiologist, a clear association is observed, but low agreement between VDG scores and BI-RADS density scores was found (weighted kappa with quadratic weights coefficient = 0.40). In general, VDG scores tend to be higher than the BI-RADS density scores. For instance, 70 studies that were scored with BI-RADS 2 obtained a VDG score of 3. The same trend was observed on 55 studies that were scored with BI-RADS 3, which obtained a VDG of 4. One should note that the VDG thresholds were set based on a US radiologist's assessment of BI-RADS density. The low agreement and the perceived overestimation might be caused by the fact that the BI-RADS scoring in this work was done by an European radiologist. BI-RADS density grades have been suggested to be underestimated according to EU standards when compared to US radiologist⁹⁶. However, further research is still required to investigate this effect as only a single radiologist participated in the presented study.

Regarding the validation process, it was a limitation of our study that we had to exclude cases without reliable breast MRI fibroglandular tissue segmentation. How-

ever, we do not think this influenced our results because the causes for rejecting MRI cases were mostly not related to breast composition. Rejected cases were distributed evenly for the BI-RADS categories 1, 2 and 3. A higher percentage of rejected cases was observed on BI-RADS category 4 (8 of 15). This fact is explained by the difficulty of automatically segmenting fibroglandular tissue in breasts with high density in MRI. One could think that the exclusion of these BI-RADS category 4 cases increases the correlation coefficients between FFDM and MRI measurements. However, these rejected cases had minor influence on the complete dataset (3% of the total number of studies).

In conclusion, our study shows that it is feasible to obtain accurate measurements of absolute and relative volumes of dense breast tissue from full field digital mammograms. Availability of such measurements is crucial for the development of objective breast cancer risk models and may be used in the development of personalized screening protocols.

Automated breast cancer detection in breast MRI

5

Albert Gubern-Mérida, Robert Martí, Jaime Melendez, Jakob L. Hauth, Ritse M. Mann, Nico Karssemeijer, Bram Platel

Original title: Automated localization of breast cancer in DCE-MRI

Published in: Medical Image Analysis (2015); 20(1):265–274

Abstract

Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is increasingly being used for the detection and diagnosis of breast cancer. Compared to mammography, DCE-MRI provides higher sensitivity, however its specificity is variable. Moreover, DCE-MRI data analysis is time consuming and depends on reader expertise. The aim of this work is to propose a novel automated breast cancer localization system for DCE-MRI. Such a system can be used to support radiologists in DCE-MRI analysis by marking suspicious areas. The proposed method initially corrects for motion artifacts and segments the breast. Subsequently, blob and relative enhancement voxel features are used to locate lesion candidates. Finally, a malignancy score for each lesion candidate is obtained using region-based morphological and kinetic features computed on the segmented lesion candidate. We performed experiments to compare the use of different classifiers in the region classification stage and to study the effect of motion correction in the presented system. The performance of the algorithm was assessed using free-response operating characteristic (FROC) analysis. For this purpose, a dataset of 209 DCE-MRI studies was collected. It is composed of 95 DCE-MRI studies with 105 breast cancers (55 mass-like and 50 non-mass-like malignant lesions) and 114 DCE-MRI studies from women participating in a screening program which were diagnosed to be normal. At 4 false positives per normal case, 89% of the breast cancers (91% and 86% for mass-like and non-mass-like malignant lesions, respectively) were correctly detected.

5.1 Introduction

Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) of the breast is employed for breast cancer screening in women with cumulative lifetime breast cancer risk of more than 20-25% (US and EU guideline)¹⁶. Other indications include, but are not limited to, preoperative staging, evaluation of women treated with neoadjuvant chemotherapy and problem solving in case of inconclusive findings from other modalities. Compared to mammography, which is the image modality commonly used in regular screening, breast DCE-MRI presents higher sensitivity²⁴, especially in women with dense breasts²³. However, specificity is more variable, both for screening¹⁶ and characterization²⁴ purposes, since the examination of breast DCE-MRI depends on many factors such as reader expertise and use of adequate visualization techniques. This fact can lead to a substantial amount of false positive findings. Furthermore, breast MRI analysis requires interpretation of four-dimensional DCE data, as well as correlation to multi-parametric data from other MRI imaging sequences, and is therefore a time consuming task.

In order to help overcome these limitations, dedicated workstations are currently used in clinical practice to assist the radiologist in the detection and classification of breast lesions in DCE-MRI data. These systems mainly provide an automated kinetic assessment by color-coding the intensity changes per voxel during enhancement of the breast tissue. This automated assessment aids in the interpretation of patterns of contrast enhancement (persistent, plateau and washout enhancement) across a series of MRI volumes³⁰, but human interaction is still required to identify and characterize suspicious areas, which increases the risk of misinterpreting or overlooking breast lesions^{25,26} and may cause inter- and intra-observer variability³¹.

A computer-aided detection (CAD) system that marks the most suspicious locations of the breast can aid radiologists in the analysis of breast DCE-MRI. Such a system can reduce the interpretation time of analyzing breast DCE-MRI data and, as shown when CAD has been applied to other modalities⁹⁷⁻¹⁰⁰, otherwise missed lesions could be detected.

Only a few authors have presented algorithms to automatically detect lesions in breast DCE-MRI³⁸⁻⁴⁰. The method developed by Vignati et al.³⁸ used subtracted mean intensity projection images over time, which were normalized using the contrast uptake of the mammary vessel. Renz et al.³⁹ evaluated a fully automatic CAD system that segmented lesions with a hierarchical 3D Gaussian pyramid approach. More recently, Chang et al.⁴⁰ combined kinetic and morphological features to identify focal tumor breast lesions. However, in these previous studies, results were reported for relatively small datasets that do not cover the entire spectrum of malignant breast lesions

since non-mass-like enhancing lesions, which are lesions with abnormal enhancement larger than focus without space-occupying effect²⁹, and thus without blob-like shape, are not included. Therefore, automatic lesion detection in DCE-MRI is still an open problem. The problem is also clinically relevant since recent studies showed that lesions are regularly overlooked or misinterpreted in breast cancer screening programs with MRI^{25,26}.

In this work, we propose a novel CAD system to automatically detect breast lesions in DCE-MRI. The system is a multi-stage approach that uses blob features in combination with kinetic and morphological information of the lesion in motion corrected data. The performance of the algorithm was assessed using free-response operating characteristic (FROC) analysis, using a dataset composed of 95 studies with manually annotated malignant mass-like and non-mass-like lesions and 114 studies from patients participating in a screening program for high and intermediate risk women that were evaluated as normal.

5.2 Material and methods

5.2.1 Study dataset

The dataset used in this study contained 209 T1-weighted coronal DCE-MRI studies from different patients (age: 22-81 years, mean age: 48.5 years). Of the complete dataset, 114 DCE-MRI studies were obtained from patients participating in a high-risk screening program. These MRIs were scored as either BI-RADS 1 ($n = 87$) or BI-RADS 2 ($n = 27$). For each included MRI, at least 2 years of follow up were available with no sign of breast cancer and no previous history of breast cancer or breast surgery was reported. These patients and their scans are referred to as normal patients and normal DCE-MRI in the paper.

The remaining 95 DCE-MRI studies contained visible malignant abnormalities, 105 malignant lesions in total, of which 55 were mass-like lesions and 50 were non-mass-like lesions. For this dataset, the average effective lesion radius was 10.83 mm with a standard deviation of 6.18 mm and range of 2.50 - 29.78 mm. The malignant lesions included ductal carcinoma in situ (DCIS), invasive lobular carcinoma (ILC), adenocarcinoma, invasive ductal carcinoma (IDC) and ovarian metastases. Of the total number of lesions included in this study, 26 lesions, with radius size of 9.05 ± 4.73 (mean \pm stdev), were detected in screening examinations with MRI. The remaining 79 lesions, with radius size of 11.41 ± 6.51 , were detected in diagnostic MRI exams performed in clinical routine, for instance after detection of breast cancer in mammograms or in symptomatic patients. The distribution of breast cancer lesion types is shown in Table 5.1.

Table 5.1: Distribution of breast cancer lesion types included in the study dataset

	Mass-like	Non-mass like
DCIS	6	22
ILC	11	12
Adenocarcinoma	8	3
IDC	29	12
Ovarian metastasis	1	1
Total	55	50

Breast DCE-MRI image acquisitions were performed on either a 1.5 or 3 Tesla Siemens scanner (Magnetom Avanto, Magnetom Sonata, Magnetom Symphony and Magnetom Trio), with a dedicated breast coil (CP Breast Array, Siemens, Erlangen). All acquisitions were collected from 2003 to 2012. Pixel spacing (from 0.586 mm to 1.33 mm), coronal slice thickness (from 1 mm to 1.5 mm) and 4D volume size (left-to-right: [256 - 512] voxels, superior-to-inferior: [128 - 256] voxels, posterior-to-anterior [72 - 160] voxels, and time points: [5 - 6]) differed among acquisitions. The first time point (T_0) was acquired before intravenous agent injection. As indicated by the breast MRI guidelines from the European Society of Breast Imaging¹⁶, the first post contrast time point was acquired around the 90th second after intravenous agent injection and consecutive time points were acquired every 100 seconds approximately. The contrast medium was administered at a dose of 0.1 mmol/kg (Medrad, Warrendale, PA) using a power injector (Medrad, Warrendale, PA) at a flow rate of 2.5 mL/s, followed by a saline flush.

All lesions were retrospectively annotated by a trained medical doctor experienced in breast DCE-MRI. The annotated lesions' location, size and their classification into mass-like or non-mass-like lesion type according to the predominant feature were verified on the basis of the radiological reports. In case of doubt, a breast radiologist was consulted. The annotations were performed in an in-house developed dedicated breast DCE-MRI annotation workstation⁷¹ that provided the visualization of DCE-MRI time sequences, subtraction images and relative enhancement images. A semi-automatic segmentation algorithm was used to obtain accurate and consistent breast lesion segmentations. This method is a modified version of the smart opening algorithm initially developed for nodule and lesion segmentation in CT images¹⁰¹. Smart opening is a hybrid segmentation approach that combines threshold-based with model-based morphological processing. It has recently also been used in other studies for lesion segmentation in breast DCE-MRI¹⁰². Segmentations were made on the subtraction volume of

the first post-contrast acquisition and the pre-contrast acquisition. A subtraction volume refers to the subtraction of image intensities between two volumes. The annotator had to provide a seed point roughly in the center of the lesion for the segmentation. More than one seed point was required to annotate large lesions (mainly non-mass-like lesions). Figure 5.1 shows a set of lesion segmentation examples.

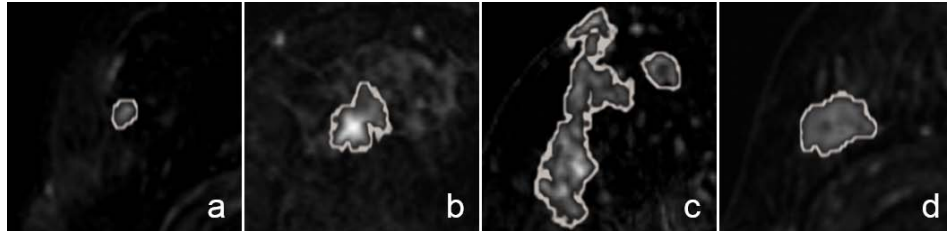


Figure 5.1: Semi-automatic lesion segmentation examples. Lesion types are (a-b) mass-like DCIS, (c) non-mass-like ILC and (d) mass-like adenocarcinoma.

5.2.2 Automatic breast lesion detection in DCE-MRI

A general overview of the automatic breast lesion detection method is shown in Fig. 5.2. The processing algorithm starts by correcting for motion artifacts and segmenting the breast. Subsequently, an initial stage to locate lesion candidates by combining blob and relative enhancement voxel features is performed on the breast area. Finally, false positive reduction is applied in a second stage by incorporating morphological and kinetic features from the candidate region. These steps are consecutively explained below.

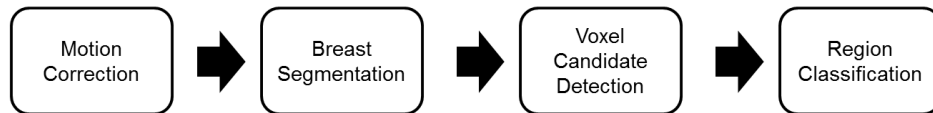


Figure 5.2: Overview of the presented automatic lesion detection framework in breast DCE-MRI.

Motion correction

Motion correction is applied to remove artifacts from the DCE-MRI data. These artifacts are caused by movement of the patient during the image acquisition process and could influence the lesion kinetic characteristics. Images at post-contrast time points ($T_1 - T_n$) are mapped to the pre-contrast acquisition T_0 by using a registration algorithm that combines rigid and non-rigid B-Splines transforms in a multi-resolution scheme¹⁰³. Stochastic gradient descent optimizer and mutual information similarity measure, as signal intensity varies along time, are used. B-Splines grid spacing was set to 160, 80 and 40 mm for each of the 3 resolutions respectively. A coarse spacing was set

in order to avoid local volume changes in the region of the enhancing lesion¹⁰⁴. Elastix⁵⁶ was used for the implementation. Figure 5.3 shows an example of a subtraction image computed before and after applying motion correction.

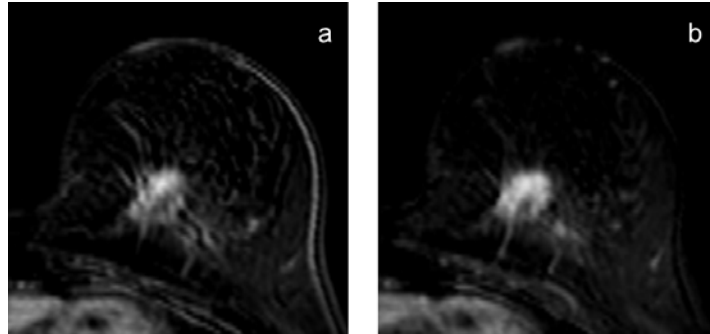


Figure 5.3: Comparison between subtraction images (a) without and (b) with motion correction for an axial slice with a visible IDC. The same window and level were used to display these images.

Breast segmentation

The breast is automatically segmented to focus the detection only on the breast tissue, excluding the chest, the arms and the pectoral muscle, and therefore reduce false positive findings. A breast MRI segmentation method that captures the anatomic variation of the pectoral muscle and chest wall region is applied to the pre-contrast DCE-MRI volume T_0 . The method is based on our previously reported work¹⁰⁵. It uses spatial information provided by probabilistic atlases. A probabilistic atlas is a volume that contains the complete spatial distribution of probabilities for a voxel to belong to one or more organs. In this work, three probabilistic atlases are used to have a good representation of differences in women with small, medium and large breasts. The size of the breast is measured as the distance along the anterior-posterior axis between the anterior of the breast and the coronal plane through the sternum. This distance is noted as d_{AP} (see Fig. 5.4(a)). The large probabilistic atlas is chosen to segment patients with d_{AP} larger than 95 mm, the medium atlas for patients with d_{AP} between 66 and 95 mm, and the small atlas for patients with d_{AP} smaller than 66 mm. Figure 5.4(b) shows an example of the segmentation obtained by the automatic breast segmentation approach.

Voxel candidate detection

After segmenting the breast, a voxel candidate detection is applied to select enhancing voxel candidates for further processing (see Sec. 5.2.2). Since signal intensity enhancement after contrast agent administration is visible in all breast lesions apart from cysts, these enhancing voxels are suspicious to belong to a breast lesion. A classifier trained

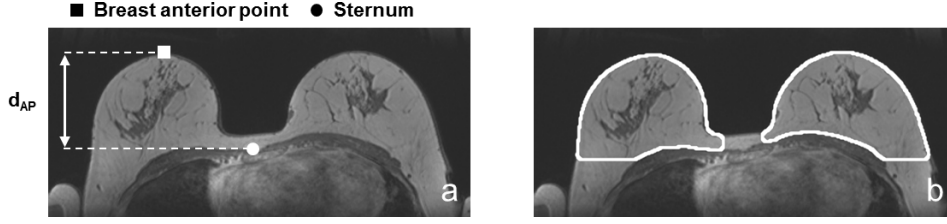


Figure 5.4: Automatic breast segmentation: axial slice of pre-contrast DCE-MRI volume T_0 with (a) breast size measured as d_{AP} and (b) its breast segmentation.

with labeled voxel data calculates the likelihood of abnormality using voxel features computed from the relative signal enhancement (R_t). The relative signal enhancement at a specific time point t after contrast agent injection is defined as

$$R_t = \frac{T_t - T_0}{T_0}. \quad (5.1)$$

As the first post-contrast time point T_1 is the advocated time point for lesion detection and assessment of morphology on breast DCE-MRI data acquired following the European guidelines¹⁶, R_1 was used in this work for the computation of features.

The first features are smoothed R_1 values at different scales, which are defined as $\bar{R}_1(\sigma_r) = R_1 \otimes G(\sigma_r)$, where \otimes represents the convolution operator and $G(\sigma_r)$ is a Gaussian smoothing filter at scale σ_r . With the computation of $\bar{R}_1(\sigma_r)$, artifacts produced by noise and local deformations not corrected in the motion correction step are reduced.

The remaining features derive from two different blob measures. These blob features are defined to characterize the shape of bright blob-like structures. These features were chosen because mass-like breast lesions and most enhancing areas of non-mass-like lesions appear as bright blobs in relative signal enhancement volumes. The first blob measure $B_1(\sigma_b)$, where σ_b represents the size of the blob, is based on the Laplacian of Gaussian. The second blob measure $B_2(\sigma_b)$, which is extensively described in Li et al.¹⁰⁶, uses eigenvalues from the Hessian matrix to characterize blobs. The eigenvalues are λ_1 , λ_2 and λ_3 and satisfy $|\lambda_1| \geq |\lambda_2| \geq |\lambda_3|$. The bright blob enhancement measure used in this work is computed as follows:

$$B_2(\lambda_1, \lambda_2, \lambda_3) = \begin{cases} |\lambda_3|^2 / |\lambda_1| & \text{if } \lambda_1 < 0, \lambda_2 < 0, \lambda_3 < 0, \\ 0 & \text{otherwise.} \end{cases} \quad (5.2)$$

In total, 16 features are used in this stage: 10 $\bar{R}_1(\sigma_r)$ at different σ_r scales exponentially ranging from 0 to 10 mm, and 6 blob features from computing $B_1(\sigma_b)$ and $B_2(\sigma_b)$ at $\sigma_b = 3, 10$ and 17 mm, respectively. The σ_r and σ_b values were selected to cover a large range of lesion sizes. An example of blob feature responses can be seen in Fig. 5.5.

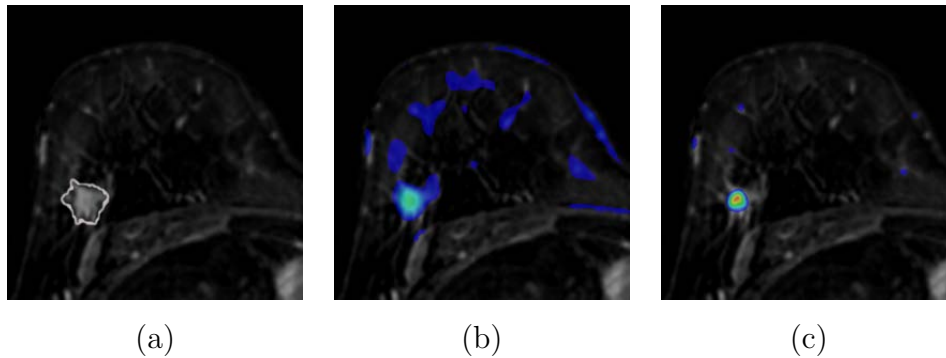


Figure 5.5: Response of the blob features (b) $B_1(\sigma_b)$ and (c) $B_2(\sigma_b)$ with $\sigma_b = 3$ mm in an axial subtracted slice containing an invasive lobular carcinoma.

The random forests classifier¹⁰⁷ was used in the voxel candidate detection stage to compute the likelihood of abnormality map. We chose this classifier as it is considered a state-of-the-art classifier and it incorporates a type of greedy feature selection as part of its learning algorithm. The maximum number of trees and the maximum tree depth were set to 500 and to the number of features used in this stage ($n = 16$), respectively. The abnormality likelihood given by the first stage was computed by averaging the output over all the trees.

To train the voxel candidate detection classifier to detect malignant lesions, we collected 40000 training samples with a ratio of 5 normal samples per abnormal sample. For the abnormal class, we collected samples within the lesion annotations with the highest relative signal enhancement R_1 voxel value. For the normal class, we randomly selected samples from breast voxels on normal cases with $R_1 > 0$.

Finally, to obtain the lesion candidates that are taken to the final detection stage, local maxima of the voxel likelihood map computed using the previously described features are determined by means of a spherical kernel (radius = 5 mm). Subsequently, the highest maxima within a radius of 20 mm are taken as the final candidates. We chose for this two-stage approach and not a single sweep with a 20 mm spherical kernel to reduce computation time.

Region classification

In this final step, the likelihood of malignancy of each candidate is computed to obtain the final CAD mark. Firstly, the lesion candidate locations obtained from the previous stage are used as seed points for an automatic 3-D region segmentation using smart opening¹⁰². As performed for the manual annotation of the lesions (see Sec. 5.2.1), smart opening is applied on the subtraction of the first post-contrast acquisition and the pre-contrast acquisition. Secondly, morphological and kinetic features are computed to train a classifier for false positive reduction.

To characterize lesion morphology, the following features are computed based on the segmentation of the lesion candidate (see Newell et al.³¹, Gilhuijs et al.^{108,109} and Behrens et al.⁵⁷ for exact definitions):

- Volume, in cubic centimeters.
- Margin sharpness and variance, taking inhomogeneity of the margin gradients and their variance along the margin into account to provide a measure of how well the lesion is delimited.
- Circularity, defined as the percentage of the lesion's volume covering the volume of a sphere with the same volume.
- Irregularity, describing the smoothness of the segmented lesions surface.
- Convexity, calculated by comparing the surface of the convex hull of the segmentation with the actual surface of the segmented lesion candidate.
- Radial gradient analysis, measuring the gradients of the lesion margin pointing toward the lesion center.
- Effective diameter.
- Degree of elongation.

To characterize the kinetic curve measured at each voxel, four well-established kinetic features are used¹¹⁰. These parameters are maximum relative signal enhancement (ME), time to peak (TTP), uptake rate (UR) and washout rate (WR). Their definitions can be found below:

$$\begin{aligned}
 ME &= \max(R_t : t = 0, \dots, N), \\
 TTP &= t_p, \text{ time of acquisition } p, \\
 &\quad \text{where } p = \operatorname{argmax}_t(R_t) \\
 UR &= ME/TTP \\
 WR &= \begin{cases} \frac{R_p - R_N}{t_N - t_p} & \text{if } p \neq N \\ 0 & \text{if } p = N. \end{cases} \quad (5.3)
 \end{aligned}$$

In total, 25 features are used in the final region classification step. These include the likelihood of abnormality of the previous stage, 8 morphology features, the average and the standard deviation of each kinetic parameter for the entire lesion candidate extent and the average and the standard deviation of each kinetic parameter measured

within the most suspicious region of the lesion candidate extent. The most suspicious region is defined as the 3x3x3 voxel volume that shows the strongest relative signal enhancement at the first post-contrast time point. To train the region classifier, we used region candidates from MR scans with and without annotated lesions. Manual annotations were used to assign the training region candidates to either the abnormal or normal class.

5.2.3 Experimental studies

Evaluation measures

FROC analysis was used to determine the performance of the system based on the malignancy level obtained for each mark and ground truth annotations. A mark given by the system was considered a true positive when the manual annotation of a malignant lesion was hit. When multiple marks hit the annotation of a lesion, the one with the highest malignancy score was chosen and the rest were discarded. False positive detections are computed on normal patients in order to evaluate the performance of the proposed approach when applied to a screening population. FROC analysis was carried out for all malignant lesions and for mass-like and non-mass-like lesions, individually.

In order to obtain a single performance measure, the mean sensitivity in the range between 0.1 and 4 false positives per normal case (FPs/case) on a logarithmic scale was computed as

$$S(i, j) = \frac{j}{\ln j - \ln i} \int_i^j \frac{s(f)}{f} df, \quad (5.4)$$

where i and j are points that limit the evaluated range, f is the number of false positives per normal case and $s(f)$ is the lesion sensitivity. The computed measure is proportional to the partial area under the FROC curve plotted on a logarithmic scale. Using a logarithmic scale avoids that the measure is dominated by operating points at high false positive rates.

Statistical significance of the performance difference between pairs of evaluated approaches was determined by means of bootstrapping¹¹¹. Cases were sampled with replacement from the pooled cross-validation set 5000 times. Every bootstrap sample had the same number of subjects as the study dataset. For each new sample, two FROC curves were constructed using the likelihood scores yielded by the two methods being compared. Then, the difference in mean sensitivity S in the range between 0.1 FP/case and 4 FPs/case $\Delta S(0.1, 4)$ was computed. After resampling 5000 times, 5000 values of $\Delta S(0.1, 4)$ were obtained. p -values were defined as the fraction of $\Delta S(0.1, 4)$ values that were negative or zero. Performance differences were considered to be significant if $p < 0.05/N$, with N the number of comparisons performed for Bonferroni correction.

Study design

In this study, the use of different classifiers in the region classification stage and the influence of the motion correction step in the final detection results were investigated. To obtain unbiased detection results, a 5-fold cross-validation was performed at patient level. The 209 subjects of the study dataset (114 normal and 95 with annotated malignant lesions) were randomly assigned to each fold. We took care that the number of malignant lesions was approximately the same for all the folds. For each fold, the corresponding training set was used to train the voxel classifier of the first stage and to obtain training lesion candidates after applying the aforementioned trained classifier. Subsequently, the region classifier was trained. The features of this stage were computed on the training lesion candidates. Finally, the complete system was applied to the test set corresponding to the current fold.

In the first experiment, lesion detection was performed using five different classifiers in the final region classification stage. Optimization of the parameters was carried out by means of grid search on a nested 5-fold cross-validation within the fold's training set. The evaluated classifiers were:

1. Linear discriminant analysis classifier (LDA).
2. k -nearest neighbors classifier (kNN): The grid search was performed on k values of 25, 50, 100, 150 and 175.
3. Gentleboost classifier (GB): For the gentleboost classifier¹¹², regression stumps were used as weak learners. The explored number of regression stumps were 50, 100, 500 and 1000.
4. Support vector machine classifier with radial basis function kernel (SVM): LIBSVM¹¹³ was used for its implementation. Cost C and kernel scale γ were optimized. C was varied by 2^i , with i between -4 and 4 with a step of 2, and γ was varied by 2^j , with j between -15 and 15 with steps of 5. Probability estimates were generated with LIBSVM following the approach of Wu et al.¹¹⁴.
5. Random forests classifier (RF): 50, 100, 500 and 1000 maximum number of trees were explored by the grid search optimization process. The maximum depth of the tree was set to the number of features ($n = 22$). The malignancy likelihood given by the classifier was computed by averaging the output over all the trees.

The mean sensitivity in the range between 0.1 and 1 FPs/case, $S(0.1, 1)$ was used as an optimization measure. The parameters that resulted in the highest $S(0.1, 1)$ value averaged over all inner five folds were chosen to detect lesions in the outer testing fold.

The second experiment was performed to investigate the benefit of applying motion correction to the data. Following the same nested 5-fold cross-validation procedure used in the previous experiment, the presented CAD was trained and tested on the study dataset with and without applying motion correction.

5.3 Results

Results of the first experiment can be observed in Fig. 5.6, which shows the performance of the proposed CAD system using different classifiers in the final region classification stage. Table 5.2 shows p -values for the comparisons between the classifiers. For the detection of malignant lesions, LDA, kNN, gentleboost, SVM and random forests obtained mean sensitivity values in the range between 0.1 and 4 FPs/case of 0.48, 0.60, 0.66, 0.64 and 0.72, respectively. The random forests classifier outperformed the other classifier techniques achieving an overall sensitivity of 0.55, 0.78, 0.89 and 0.95 at 0.1, 1, 4 and 7 false positives findings on normal patients, respectively. The sensitivities for MRI screen-detected lesions and for lesions detected in MRI after referral from other modalities were 0.46, 0.69, 0.81 and 0.89, and 0.58, 0.81, 0.91 and 0.96 at 0.1, 1, 4 and 7 FPs/case, respectively. Considering that 10 pairwise comparisons between different classifiers were performed, the random forests classifier obtained significantly better performance than LDA and kNN ($p = 0.001$ and $p < 0.001$, respectively). The performance of the initial voxel candidate detection stage is also presented in Fig. 5.6.

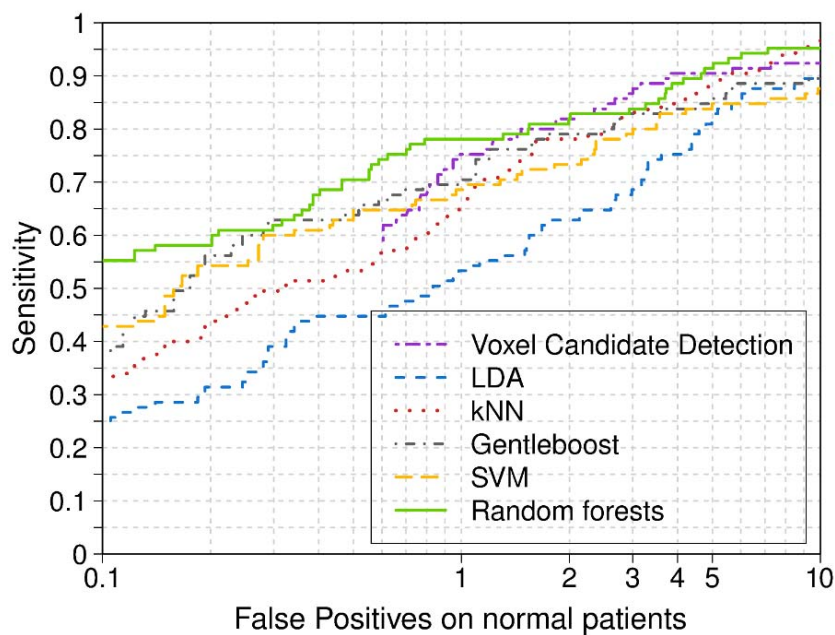


Figure 5.6: FROC curves of the presented CAD system using different classifiers in the region classification stage.

Table 5.2: p -values obtained when comparing the performance of different classifiers in the region classification stage. Significant differences are indicated with a star (*).

	LDA	kNN	GB	SVM	RF
LDA	-	0.002*	<0.001*	<0.001*	0.001*
kNN	-	-	0.067	0.042	<0.001*
GB	-	-	-	0.679	0.037
SVM	-	-	-	-	0.011
RF	-	-	-	-	-

Figure 5.7 shows some examples of lesions detected by our CAD system using random forests in the region classification stage. Examples of false positive detections in normal patients can be observed Fig. 5.8.

In the second experiment, the presented CAD system was trained and tested on the study dataset with and without applying motion correction. The random forests classifier was used in the region classification stage. The FROC curves obtained for the detection of all malignant, mass-like malignant and non-mass-like malignant lesions are shown in Fig. 5.9. The use of motion correction significantly improved the overall performance of the system to detect breast cancer ($p = 0.023$) reaching a sensitivity for the detection of mass-like and non-mass-like malignant lesions at 0.1, 1, 4 and 7 FPs/case of 0.67, 0.85, 0.91 and 0.96, and 0.42, 0.7, 0.86 and 0.92, respectively. As visible in the FROC curve of Fig. 5.9(c), motion correction has a larger impact in the detection of non-mass-like malignant lesions ($p = 0.004$). Figure 5.10 shows a DCIS that was detected at 0.38 FPs/case when motion correction was applied. Without motion correction, the lesion shown in Fig. 5.10(a) was missed by the system.

5.4 Discussion

In this work, we developed a CAD system for breast cancer in DCE-MRI that provides automated localization of malignant lesions on motion corrected data. First, relative enhancement and blob features are used to detect suspicious areas. Then, the final malignancy likelihood is obtained using region-based morphological and kinetic information computed on segmented lesion candidates. The use of different classifiers in the final region classification stage and the effect of applying motion correction were investigated. A large dataset of 209 DCE-MRI studies from different women composed of 95 DCE-MRI volumes with 105 annotated mass-like and non-mass-like malignant lesions, and 114 DCE-MRI volumes from normal screening patients was used to evaluate the algorithm when applied to a screening population.

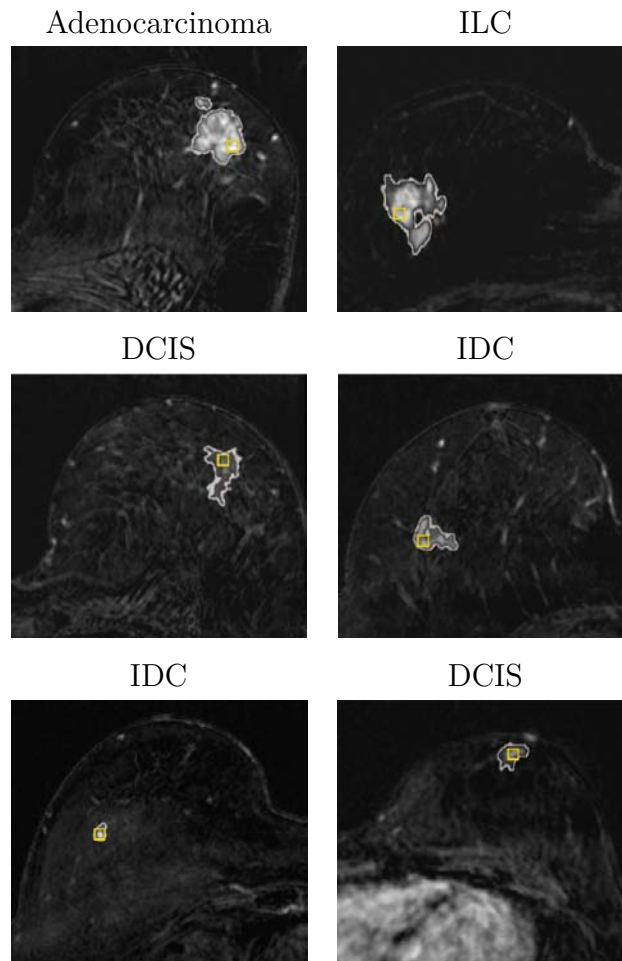


Figure 5.7: Lesions detected by the proposed CAD system on axial subtracted images at the first post-contrast time point. The squared markers represent the lesion candidates given by the proposed algorithm.

An overall sensitivity of 0.55, 0.78, 0.89 and 0.95 at 0.1, 1, 4 and 7 false positives findings on normal patients, respectively, was achieved for the detection of malignant lesions. The mean sensitivity in the range between 0.1 and 4 FPs/case was 0.72. The best results were obtained using random forests in the final region classification stage. The random forests classifier outperformed LDA, kNN, gentleboost and SVM. This improvement in performance was significant when random forests were compared to LDA and kNN. The performance of the CAD was also evaluated for the detection of mass-like and non-mass-like lesions separately. The sensitivity for the detection of mass-like malignant lesions was 0.67, 0.85, 0.91 and 0.96 was obtained at 0.1, 1, 4 and 7 FPs/case, respectively. The performance was lower for the detection of non-mass-like malignant lesions: 0.42, 0.7, 0.86 and 0.92 at 0.1, 1, 4 and 7 FPs/case. The sensitivity for the detection of lesions detected in MRI after referral from other modalities (0.58, 0.81, 0.91 and 0.96 at 0.1, 1, 4 and 7 FPs/case), as expected, was higher than the

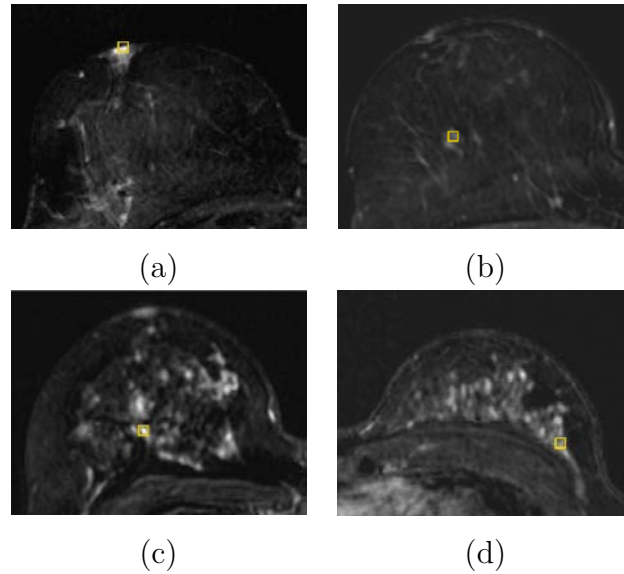
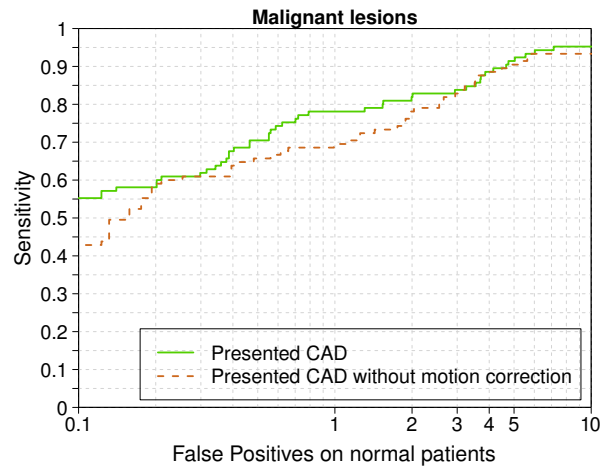


Figure 5.8: False positive detections on axial subtracted images of 4 different normal patients. The squared markers represent the lesion candidates given by the proposed CAD system.

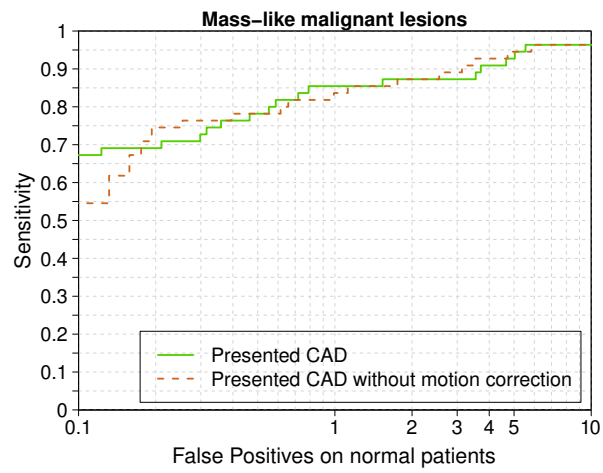
sensitivity obtained for the detection of screen-detected lesions (0.46, 0.69, 0.81 and 0.89, at 0.1, 1, 4 and 7 FPs/case). This is explained by the fact that the former includes larger lesions, which is in line with the findings in symptomatic patients¹¹⁵.

To date, only a few studies on automatic breast lesion detection in DCE-MRI are found in the literature^{38–40}. Results for the detection of malignant lesions were 0.98 at 4 FP per breast³⁸, 0.98 at 0.16 FPs/case³⁹ and 1 at 6.30 FP/case⁴⁰. Care should be taken when comparing the presented results to the results reported in these previous studies, since different datasets were used and different evaluation procedures were followed. For instance, false positive detections were not computed on normal cases from women participating in a high-risk screening program^{38–40}, but only computed false positive marks as detections in other locations than the apparent cancer in women with a lesion on the MRI scan. Note that under a screening setting, only the normal cases are relevant when assessing the false-positive detections. Furthermore, ground truth annotations were not provided in Renz et al.³⁹ and in Chang et al.⁴⁰, but their evaluation was based on the visual inspection of the findings detected by the system. In Vignati et al.³⁸ and in Renz et al.³⁹, FROC analysis was not performed. Finally, only mass-like lesions were considered in Vignati et al.³⁸ and in Renz et al.³⁹. Chang et al.⁴⁰ do not clearly describe whether non-mass-like lesions were also included in their evaluation dataset.

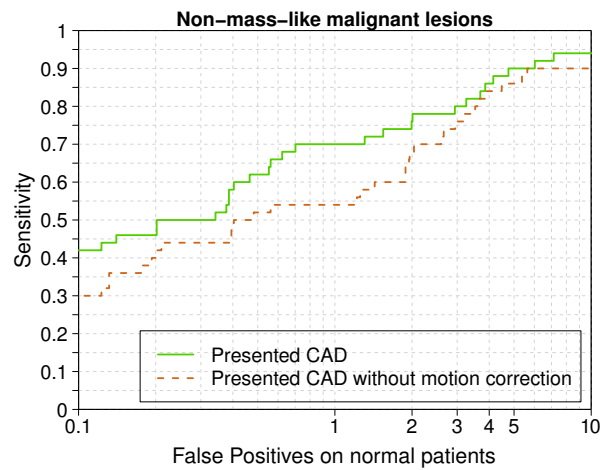
The detection of non-mass enhancement is important as a large proportion of relevant lesions present as non-mass-like enhancement. In practice, 20 to 30 % of all invasive cancers present as non-mass-like enhancement¹¹⁶. Moreover, the vast majority of DCIS lesions presents as non-mass-like enhancement. While it is known that DCIS



(a)



(b)



(c)

Figure 5.9: FROC curves yielded by our CAD system with and without motion corrected data for (a) all malignant, (b) mass-like malignant and (c) non-mass-like malignant lesions.

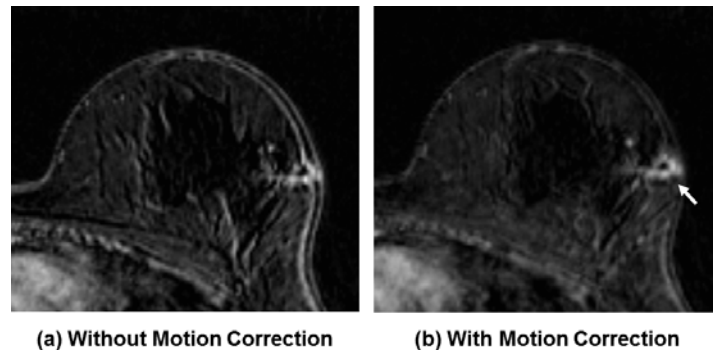


Figure 5.10: A DCIS detected at lower false positive rate by the CAD system when motion correction was applied. The same window and level were used to display these axial subtracted images. Lesion is pointed by an arrow on the motion corrected example. Due to motion, the relative enhancement of healthy breast parenchyma can appear as enhancing structures that can confuse the CAD system both in training and testing. This is illustrated in the left part of the image (a). Additionally, motion can introduce spurious edges in lesion morphology (see lesion in (a)). Both issues are reduced by motion correction (b).

is not harmful at this stage due to its noninvasive form, evidence suggests that DCIS lesions may progress to invasive carcinomas if they are left untreated¹¹⁷. MRI is particularly useful for the detection of high grade DCIS, that is known to progress faster to invasive disease than low grade DCIS¹¹⁸, and usually also presents as non-mass-like enhancement on MRI.

In this work, we also studied the effect of motion correction in terms of automatic detection of breast cancer. We showed that applying motion correction to DCE-MRI data improved the performance of our CAD system. This benefit is mainly observed for the detection of non-mass-like lesions, which have more irregular shapes. We observed that the enhancement pattern visible on non-mass-like lesions is more uniform on motion corrected data, which facilitates the segmentation of the lesion candidates and the computation of kinetics and morphological features. In our experiments, we did not observe this effect on mass-like lesions as the shape and the enhancement of this type of lesions were not strongly altered by motion artifacts. However, we believe that there is still room for improvement as a B-Splines registration algorithm with a coarse grid spacing was used in this work. The effect of using a finer grid spacing and/or other motion correction algorithms that account for smaller deformations remains to be investigated.

The first post-contrast time point was used to compute relative enhancement and blob features in the voxel candidate detection stage. The first post-contrast time point was chosen because, considering the protocol used to acquire the study dataset, this is the advocated time point for lesion detection and assessment of morphology¹⁶. One should note that protocols for performing breast DCE-MRI may substantially vary

among centers, due to the equipment being used and the clinicians' preferences. The use of a different time point might be required when applying the proposed CAD system to data acquired in other centers, or ideally, specific algorithms could be developed to estimate the contrast arrival time and select the most appropriate time point, which would not imply any modification in the proposed framework.

Regarding the evaluation of this work, it is important to note that the aim of our detection algorithm is to locate suspicious lesions in order save interpretation time and avoid overlooking errors. For this reason we used BI-RADS 1 and 2 cases as normal cases in our training and testing. These lesions were identified by our radiologists as obviously benign. All biopsy proven malignant lesions were included as true positive training and test samples. It is debatable whether biopsied benign BI-RADS 3 or higher lesions should be considered as true or false positives, since they are suspicious lesions that require further investigation. This is why we left them out both in our training set as well as in our test set. Further classification of the lesions detected by the presented system into benign or malignant could be performed using existing lesion characterization techniques^{31,110,119,120}.

Further research will focus on improving the performance of the presented algorithm at a low false positive rate. We observed that the use of only one seed for lesion segmentation may produce less accurate results, which would influence the computation of morphological and kinetic features. This effect was specially visible in non-mass-like lesions. We believe that a more accurate lesion segmentation algorithm can help to improve the overall performance of our algorithm. In order to do so, similar to the procedure followed in this work to obtain manual annotations, techniques can be developed to identify candidates belonging to the same lesion and, therefore, apply smart opening on multiple seeds. Other breast lesion segmentation techniques will also be investigated: a lesion segmentation approach based on Fuzzy C-Means¹²¹, a multichannel Markov Random field framework¹²² and a spiral-scanning method previously applied to 3D Automated Breast Ultrasound (ABUS) images for the segmentation of breast lesions¹²³. Alternatively, the use of other breast segmentation algorithms^{41,42,64,124} will be studied as two lesions were missed due to breast segmentation errors.

5.5 Conclusion

In this paper we have presented a fully automatic algorithm for the detection of breast cancer in DCE-MRI. We introduced blob features in an initial voxel candidate detection stage. Morphological and kinetic features are employed by a second step to reduce false positives. Evaluation on a dataset of 209 DCE-MRI cases, which contained 114 DCE-MRI studies representative for normal patients participating in a screening program,

showed that our approach yields high sensitivity for the detection of malignant lesions at an acceptable number of false positives: 0.55, 0.78, 0.89 and 0.95 at 0.1, 1, 4 and 7 false positives findings on normal patients, respectively. While the appropriate operative point has to be determined experimentally, we are confident that the presented algorithm has the potential to support radiologists during the analysis of DCE-MRI data by automatically prompting suspicious areas.

Automated detection of breast cancer in prior
false-negative screening MRI

6

6.1 Introduction

Breast Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI) has been used for almost 15 years^{17,125–128} as a screening imaging modality in patients with high risk for developing breast cancer. DCE-MRI has shown to be more sensitive to detect breast cancer than mammography^{17–22}, especially in women with dense breasts⁵⁸. Nowadays, breast DCE-MRI is recommended in screening programs for women with cumulative lifetime breast cancer risk of more than 20-25% (US and EU guidelines)^{15,16}. Screening allows the retrospective analysis of observer errors. Retrospective studies in mammography^{100,129–133} revealed that observer errors were frequent. Based on the results of these studies, it could be estimated that 20%-30% of cancers could be detected earlier in screening without an increase in the recall rate to an unacceptable level¹³⁴.

Only a few similar but smaller studies into observer error have been performed in breast MRI^{25,26}. These studies investigated the causes of false-negatives results on prior DCE-MRI studies in patients who developed breast cancer as revealed on a follow-up or incident round of screening MRI^{25,26}. In the retrospective evaluation of Pages et al.²⁵, 58 pairs of prior and diagnostic MR imaging studies (60 cancers) were assessed. Of these 60 cancers, the authors reported potential observer error in 28 (47%) cancers. Similarly, Yamaguchi et al.²⁶ reviewed 16 incident breast cancers in 15 patients screened with MRI. Of these 16 breast cancers, 9 (56%) were identifiable on the prior MRI. Main causes for overlooked or misinterpreted lesions were small lesion size, extensive background enhancement and malignant lesions with smooth margins, which is typically found in benign diseases.

These numbers indicate the importance of the development of additional tools to aid radiologists analyzing DCE-MRI. Although it has not been reported, it is evident that fatigue or lack of experience while analyzing 4-dimensional data are also other factors for observer error. A computer-aided detection (CAD) system that prompts suspicious regions could draw the attention of the radiologist to a tumor that might be otherwise overlooked, or to an abnormal region in the image that would require careful interpretation. Computer-aided detection systems are currently being developed for this purpose^{38–40}.

The aim of this study is to evaluate the potential use of a CAD system to detect breast cancer which was overlooked or misinterpreted in a breast screening program with MRI. To this purpose, we applied the CAD system that we have developed in this thesis (see Chapter 5) on prior DCE-MRI scans with cancer lesions that were prospectively detected on an incident round of screening.

6.2 Materials and methods

6.2.1 Case selection

This retrospective study was approved by our institutional board and the requirement for informed consent was waived. A cross computer search of our MR imaging records and our pathology records identified, between January 2003 and December 2013, a total of 573 patients with invasive cancers and 78 with in situ carcinomas. In 42 of these patients (36 with invasive, 5 with in situ carcinomas and 1 with both), there was a prior DCE-MRI examination performed between 6 - 24 months before the one where the cancer was diagnosed. The mean time interval between prior and diagnostic DCE-MRI studies was 11.9 ± 3.31 months. We only considered prior DCE-MRI studies with a Breast Imaging Reporting and Data System (BI-RADS) score of 1, 2, or 3. We also included a cancer which was scored with BI-RADS 0 on the prior MRI but BI-RADS 2 was given after an additional 3D ultrasound that was performed on the same day. In total, 44 tumors that constituted the study group were present in these 42 patients. All included patients were women between the ages of 33 and 78 years (mean age 50 ± 9.71 years). The indications for performing breast DCE-MRI examination included high risk screening due to prior breast cancer (10 patients), positive family history of breast cancer (30 patients, 22 patients with BRCA mutations), estrogen therapy (1 patient) or solving problems after conventional imaging (1 patient).

6.2.2 DCE-MRI imaging technique and retrospective interpretation

The DCE-MRI examinations in this study were performed on 1.5 or 3T Siemens scanner (Magnetom Avanto, Magnetom Sonata, Magnetom Symphony or Magnetom Trio) by using a dedicated bilateral breast surface coil. Patients were placed in prone position. A transverse or coronal three-dimensional T1-weighted gradient-echo (GRE) dynamic sequence was performed before contrast agent administration followed by 4 or 5 post contrast sequences. Pixel spacing (from 0.664 mm to 1.5 mm), slice thickness (from 1 mm to 1.5 mm), matrix (256×128 , 384×192 , 448×381 or 512×96 pixels), echo time (from 1.71 msec to 4.76 msec), repetition time (from 4.56 msec to 8.41 msec) and flip angle (from 10° to 25°) differed among acquisitions. The first time point was acquired before intravenous contrast agent injection. The first post contrast acquisition started within the first 2 minutes after intravenous contrast agent injection and consecutive time points were acquired every 100 seconds approximately¹⁶. The contrast medium was administered at a dose of 15 mL or 30 mL using a power injector (Medrad, Warrendale, PA) at a flow rate of 2.5 mL/s, followed by a saline flush.

The DCE-MRI studies were retrospectively evaluated by a trained technical physician with expertise in breast MR imaging in an in-house developed dedicated breast

DCE-MRI annotation workstation⁷¹. The workstation provided the visualization of subtracted images, their maximum intensity projection and T1-weighted images for all the time points of prior and diagnostic DCE-MRI studies. Figure 6.1 shows an example. For each lesion of the diagnostic MR examinations, a manual segmentation was made and the morphological and enhancement characteristics using the BI-RADS MR-lexicon²⁹ were given. Simultaneously, the prior MR examination was analyzed. If the lesion was detected, its level of visibility in the prior MRI was scored as “minimal sign” or “visible” and the manual segmentation and the BI-RADS lexicon report were also given. Annotations and BI-RADS assessment were performed using motion corrected data (see Sec. 5.2.2).

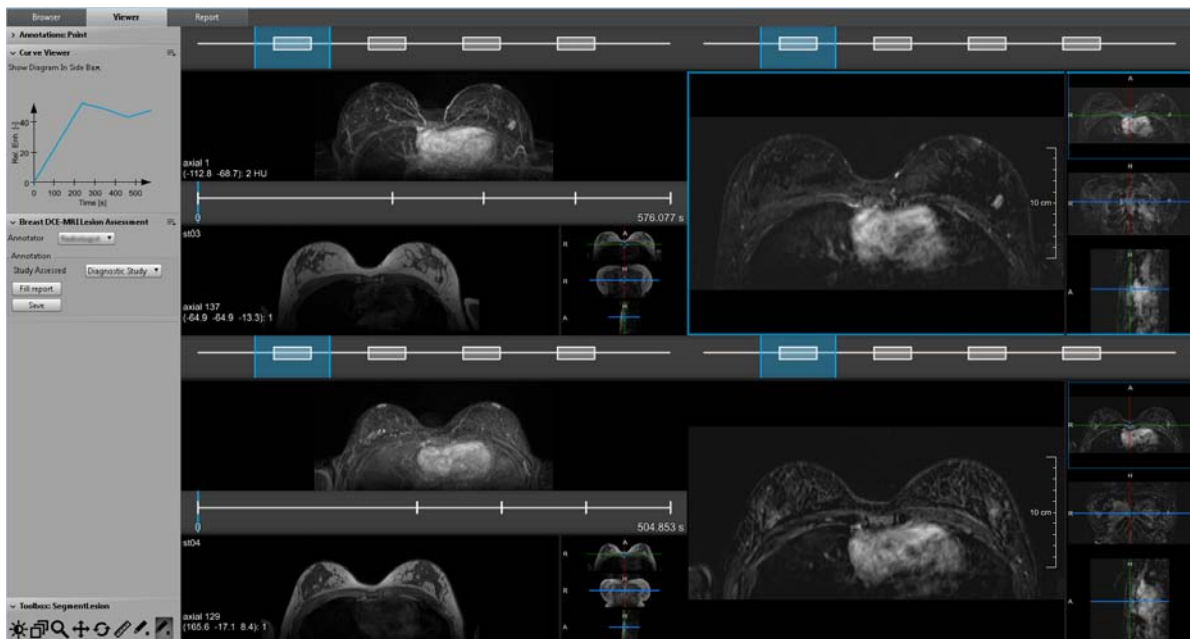


Figure 6.1: Workstation used in this study for lesion annotation and BI-RADS lexicon reporting in diagnostic (upper row) and prior (lower row) DCE-MRI studies.

Of the 44 retrospectively reviewed lesions, 13 were annotated as not visible, 15 as minimally visible and 16 as visible in the prior DCE-MRI. Figure 6.2 shows an example of each category.

6.2.3 Application of CAD on false-negative studies

In this work, the CAD system described in chapter 5 was applied to the 42 prior DCE-MRI scans of the study dataset. In short, the CAD initially corrects for motion artifacts and segments the breast. Subsequently, lesions candidates are detected using relative enhancement and texture features that include blob descriptors. The final classification is performed using region-based morphological and kinetic features computed on segmented lesion candidates. Scans of the patients included in this study were not used to

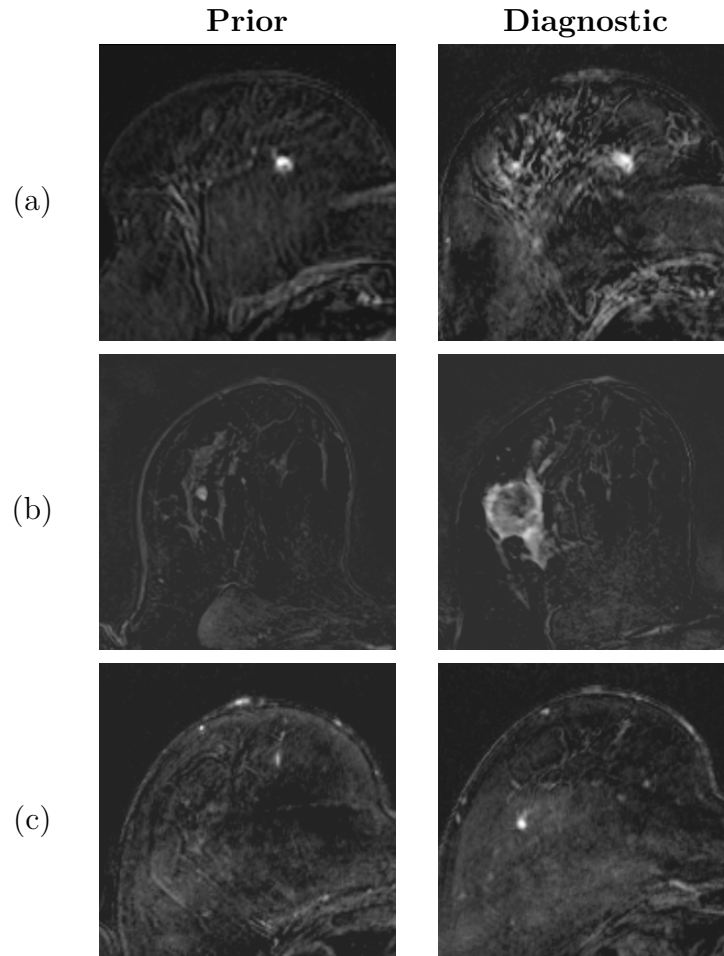


Figure 6.2: Subtracted images with breast cancer lesions which are (a) visible, (b) minimally visible and (c) not visible on the prior DCE-MRI study. BI-RADS scores were 1 and 4, 2 and 5, and 1 and 3 for prior and diagnostic of (a), (b) and (c), respectively

train the CAD system.

Of the 42 prior DCE-MRI studies, CAD markers were obtained on 41 cases (31 lesions). The CAD system failed to process one case which had very deviating anatomical characteristics due to thorax deformation and scoliosis (see Fig. 6.3). During MR acquisition, the disease caused the patient lying in a different position than the assumed by the CAD system to process the image resulting in a failure in the breast tissue segmentation. This case was excluded from the evaluation.

We studied the detection performance of visible or minimally visible lesions in false-negative DCE-MRI studies. A false-negative DCE-MRI study was defined as a prior DCE-MRI examination with breast cancer visibility rated as minimal sign or visible. The detection performance was evaluated using Free-Response Operating Characteristics (FROC) analysis. A finding detected by the system was considered a true positive when its center was located inside the manual lesion annotation. False-positive detections were computed on normal breasts. In this study, these were breasts without

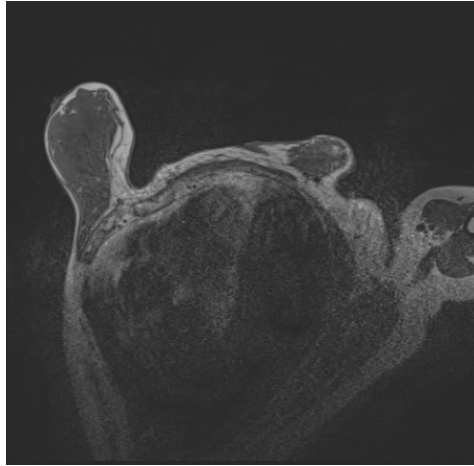


Figure 6.3: Prior DCE-MRI study of a patient with post-polio scoliosis and thorax deformation.

reported malignant disease in prior and diagnostic DCE-MRI scans. Mean sensitivity and 95% confidence interval (CI) from 0 to 10 false-positives per normal breast (FPs/breast), with steps of 0.05, were determined using bootstrapping¹¹¹ to build the FROC curves. Cases were sampled with replacement from the pooled cross-validation set 5000 times. Every bootstrap sample had the same number of subjects as the study dataset. For each new sample, the FROC curve was constructed using the likelihood scores yielded by the CAD system. Then, the sensitivity at the given FPs/breast point was computed. After resampling 5000 times, 5000 values of sensitivity at the given FPs/breast point were obtained.

6.3 Results

Figure 6.4 shows the performance of the CAD system when applied to false-negative DCE-MRI studies. FROC curves for the detection of visible and minimally visible and all lesions are illustrated. Mean sensitivity of the CAD system measured with bootstrapping was 0.63 (95% CI = 0.36 - 0.88) and 0.19 (95% CI = 0.00 - 0.43) at 2 FPs/breast for visible and minimally visible lesions, respectively. Mean sensitivity was 0.43 (95% CI = 0.24 - 0.61) for the detection of all these lesions combined. At 1 FPs/breast, mean sensitivity was 0.44 (95% CI = 0.15 - 0.75), 0.04 (95% CI = 0.00 - 0.31) and 0.25 (95% CI = 0.07 - 0.50) for visible, minimally visible and all lesions, respectively. Figure 6.5 shows some examples of lesions which were retrospectively reported as minimally visible or visible in the prior DCE-MRI study and were detected by the CAD system.

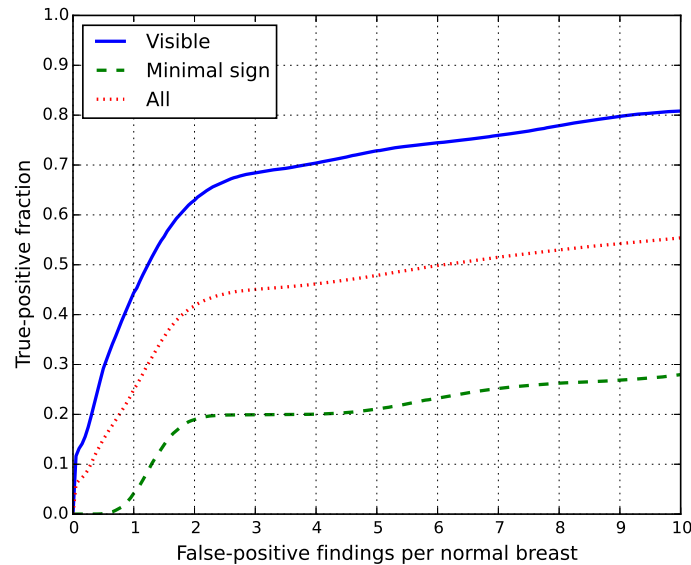


Figure 6.4: FROC curves of the CAD performance when applied to prior DCE-MRI studies with minimal sign and visible breast cancers.

6.4 Discussion

In this study, we evaluated the performance of an automated CAD system when applied to screening DCE-MRI studies with misinterpreted or overlooked breast cancer. A dataset of 41 pairs of prior and diagnostic DCE-MRI studies with detected breast cancer was used. In a retrospective review of these cases, we found that 16 and 15 breast carcinomas were visible and minimally visible, respectively, in the prior DCE-MRI scan. Results indicate that 63% of the visible and 19% of the minimally visible lesions were detected by our CAD system at 2 false-positives per normal breast.

Although it is not clear which number of false-positives per study is required for a computer-aided detection system to be effective, these numbers indicate that an important amount of overlooked or misinterpreted cancers could have been detected if a CAD system such as ours would have marked early malignant signs to the screening radiologists. Large scale experiments need to be set up to determine an appropriate setting. However, we believe that 2 false positive per breast (4 false positives per study or volume) could be acceptable as we are dealing with volumetric data that represents both breasts in a single volume. Note that commercial mammography and chest x-ray CAD systems have about 1-3 marks per case^{135,136} and 0.5-2 marks per image¹³⁷, respectively.

Further research needs to be conducted to find out what the causes were for the false-negative judgments of the radiologist who interpreted the original scans. Obviously, this is difficult to do retrospectively. In previous studies, a panel of experts was used to classify false-negative findings in prior DCE-MRI studies with cancer revealed

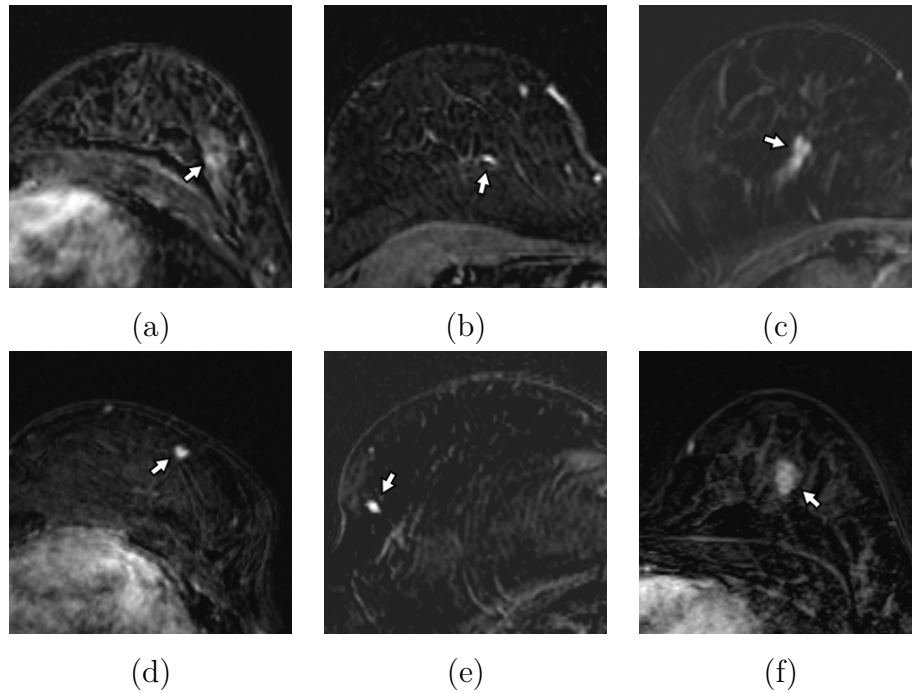


Figure 6.5: Subtraction images with breast cancer lesions (a-b) minimally visible and (c-f) visible on prior DCE-MRI studies. These lesions were automatically detected by the CAD system.

on follow-up DCE-MRI. Lesion types were classified according to BI-RADS lexicon MRI²⁹ and diagnostic errors were categorized as mismanagement, misinterpretation or oversight^{25,26}. In future studies we will also investigate the performance of the CAD system according to lesion types and/or diagnostic errors.

It is important to note that the overall performance of the CAD system might improve by specifically training on misinterpreted or overlooked lesions. These cases, which can be considered as difficult, were not included in the training phase of the CAD system used in this study.

In conclusion, a computer-aided detection system has the potential to automatically detect lesions misinterpreted or overlooked in a screening program with MRI for breast cancer. The integration in clinical practice of such a system may aid the radiologist to avoid reading errors.

Summary and discussion

Summary and discussion

Breast cancer is the most common cause of death among women, worldwide¹. Since early detection increases the chance of survival, many countries have introduced breast cancer screening programs. In regular breast cancer screening programs, asymptomatic women are periodically invited for a mammographic examination. However, mammography has its limitations as tumors can get obscured by the presence of dense tissue^{10,58,76}. Density has been also identified as a risk factor for developing breast cancer¹⁰. Other risk factors that are associated with increased risk are family history of breast cancer and the presence of a germ-line mutation of the *BRCA1* or *BRCA2* gene^{13,14}.

An approach to overcome these shortcomings and optimize breast cancer screening is the introduction of personalized screening¹¹. In personalized breast screening programs, women with high risk for developing breast cancer are screened with a complementary modality. Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) has shown potential as a breast cancer screening tool for this high risk population. Compared to mammography, DCE-MRI has higher sensitivity¹⁷⁻²², especially in women with dense breasts²³. Moreover, MRI is particularly useful for the detection of high grade ductal carcinoma in situ (DCIS), that is known to progress faster to invasive disease than low grade DCIS¹¹⁸. However, the analysis of breast MRI data requires interpretation of four-dimensional DCE data, as well as correlation to multi-parametric data from other MRI imaging sequences, and is therefore a time consuming task. Furthermore, recent studies reported potential observer errors due to misinterpretation or oversight of breast cancer lesions that were visible on a follow-up or incident round of screening MRI^{25,26}.

This thesis focuses on the automated analysis of breast MRI since automated image analysis techniques and computer-aided detection systems are required to improve the clinical workflow and aid radiologist for reading and interpreting MRI images.

Chapter 2 reports on the development of two fully automated methods for to segmentation of the pectoral muscle in breast MRI. The identification of the pectoral muscle is an important step in methods for automatic breast cancer assessment. Two possible applications are the delineation of the boundary between body and breast, which is the most challenging task to segment the breast due to large shape variations across patients, and the use of pectoral muscle as a reference tissue for pharmacokinetic modeling. The proposed segmentation methods take into account the 3-dimensionality nature of breast MRI and are based on atlas-based methodology: a probabilistic and a multi-atlas approach. Both methods were evaluated on the same dataset. The multi-atlas framework slightly outperformed the probabilistic (mean DSC values of 0.74 ± 0.06

and 0.72 ± 0.09 , respectively) but the computation time was 14 times larger.

The main limitation of the probabilistic atlas approach is the use of a single reference atlas. When the image being segmented differs considerably from the reference atlas and the registration can not compensate the differences in shape, the final segmentation becomes affected with slightly poorer results. In order to take advantage of the larger number of atlases available in the multi-atlas approach without considerably increasing the computation time of the probabilistic atlas framework, a multi-probabilistic atlas approach was proposed. This multi-probabilistic atlas framework, which has been applied in the studies described in chapters 4 and 5, uses three probabilistic atlases which were built to represent three different populations considering the size of the breast. Fifty-three manually segmented DCE-MRI studies were used for atlas construction. Although a substantial improvement in segmentation performance was observed using the multi-probabilistic atlas algorithm compared to using a single probabilistic atlas, poorer quality segmentation results due to large difference in shape between target image and the selected reference atlas were also found. Results suggest that the increase of number of atlases to cover a larger range of shape variability among patients has the potential to produce more consistent results. Moreover, other shape differences among patients and patient information, such as weight, length, and age, could be considered to construct the atlases of each represented women population.

In chapter 3, an automated method to segment the breast and estimate breast density in breast MRI is described. Other image processing techniques that are required to perform this task, such as image normalization, bias field correction and sternum landmark detection, are also explained. The breast is segmented by automatically identifying the breast-body and the breast-background boundaries. The breast-body boundary segmentation is achieved using the probabilistic atlas approach described in chapter 2. Two-dimensional region growing is applied to segment the breast-background boundary. Subsequently, the breast volume of each breast is redefined using curvature and landmark detection to exclude fatty tissue areas that are not part of the breast and are not visible in mammographic scans. By doing so, breast density measurements computed on MRI are comparable to breast density estimates obtained from mammography. Finally, breast density is independently computed for each breast using Expectation-Maximization algorithm and skin-fold removal. Results show high agreement between automatic breast and breast density segmentations and manual segmentations obtained from experts. For breast segmentation, we achieved a mean DSC value of 0.94 ± 0.04 and a mean total overlap of 0.96 ± 0.02 . For the automatic fibroglandular tissue segmentation, we obtained a mean DSC value of 0.80 ± 0.13 .

A slightly modified version of the developed breast segmentation and breast density estimation method is used in chapter 4 to validate Volpara 1.4.3 (Mātakina, Welling-

ton, New Zealand), which is a commercially available method for assessing volumetric breast density on FFDM. Volumetric estimates obtained with Volpara from FFDM images were compared to volume estimates obtained from corresponding breast MRI data of the same patient. To that purpose, our developed method was applied to a dataset of 250 MRI scans. A radiologist with expertise in breast imaging carefully reviewed all slices of the segmentations and approved 186 (74.4%) MRI studies with segmentations to be suitable for the use as a reference standard for validation of FFDM density measurements. Volumetric breast density and breast tissue volume values obtained with Volpara present high correlation when compared to MRI measurements (correlation values per breast and per study of 0.91 and 0.93, respectively). This finding demonstrates that it is feasible to obtain accurate measurements of absolute and relative volumes of dense breast tissue from full-field digital mammograms. Availability of such measurements is crucial for the development of objective breast cancer risk models and may be used for stratification of women into personalized breast cancer screening protocols.

Although satisfactory results were obtained with the segmentation methods developed in this thesis, for demanding applications, breast segmentation should be further improved. The main reasons for poor performance were, in addition to the previously explained issues related to atlas segmentation, images with small field of view where the breast was not fully covered, and the presence of artifacts and bias field after correction. As it will be discussed later in this chapter, some of these errors could be solved by the implementation of automated image quality assessment systems that guarantee a minimum image quality criteria. Alternatively, intra- and inter-patient signal intensity variability in DCE-MRI could be improved. Normalization methods based on separating the image into different frequency bands¹³⁸, the development of dedicated breast MRI bias field correction algorithms and finding the best parameters of existing bias field algorithms⁶⁷ for their application to breast MRI are some alternatives.

Chapter 5 provides the description of a Computer-Aided Detection (CAD) system to automatically detect breast cancer in DCE-MRI. The CAD system corrects for motion artifacts and automatically segments the breast. Subsequently, lesion candidates are detected using relative enhancement and texture features that include blob descriptors. The final classification is performed using region-based morphological and kinetic features computed on segmented lesions candidates. The performance of the system for the detection of mass-like and non-mass-like lesions is evaluated. Breast cancer is detected at high sensitivity with an acceptable number of false-positive findings: 0.55, 0.78, 0.89 and 0.95 at 0.1, 1, 4 and 7 false-positives findings on normal patients. We also observed that motion correction is needed to achieve the best CAD system performance. Overall, the obtained results are promising to support radiologists during the analysis of DCE-MRI data by automatically prompting suspicious areas.

In order to improve the developed CAD system, future work can be focused on achieving higher performance and decrease the number of false-positives. The second step of the described CAD system was used to differentiate between breast lesions and normal enhancing tissue by computing region-based morphological and kinetic features on the segmented candidate lesion. We observed that the lesion segmentation algorithm used for this task produced less accurate results in larger irregular lesions, which influences the computation of morphological and kinetic features. We believe that a more accurate lesion segmentation algorithm can help to improve the overall performance of our algorithm. Other lesions segmentation algorithms can be investigated, such as a lesion segmentation approach based on Fuzzy C-Means¹²¹ and a spiral-scanning method previously applied to 3D Automated Breast Ultrasound (ABUS) images for the segmentation of breast lesions¹²³. Furthermore, the use of pharmacokinetic modeling parameters to characterize breast lesions can be investigated¹³⁹. In the experiments described in chapter 5, we also found that 2 out of 105 lesions were missed due to breast segmentation errors. The CAD system excludes from the analysis all the voxels located outside the breast segmentation. We believe that the breast segmentation algorithm can be improved with the points described above. Alternatively, the use of other breast segmentation algorithms^{41,42} can be considered.

In chapter 6, we studied the potential of our CAD system to detect breast cancer which was overlooked or misinterpreted in a breast screening program with MRI. A dataset of 41 pairs DCE-MRI studies was collected. Each pair was composed of a prior DCE-MRI and an incident or follow-up DCE-MRI study where the breast cancer was diagnosed. In a retrospective review of these cases, we found that 16 breast cancers were visible and 15 were minimally visible in the prior DCE-MRI scan. Our CAD system detected 63% of the visible and 19% of the minimally visible lesions at 2 false-positives per breast. These results suggest that the integration in clinical practice of such a system may aid the radiologist to avoid reading errors.

Finally, it should be noted that all the methods described in this thesis have been developed and evaluated using data collected in only one clinical center. Therefore, robustness against differences in population and image acquisition systems have not been investigated. However, we are confident that, with minor adjustments, the proposed methods are suitable for application to datasets obtained in different institutions and under different conditions, since our datasets contained images obtained on different scanners with different clinical imaging protocols. Furthermore, since these methods have been developed taking the anatomy of the women into account, we believe that they can also be applied to ultrafast dynamic breast MRI sequences that are being developed^{102,140}. This new imaging sequence allow much shorter and more cost-effective procedures, which is beneficial especially for screening.

To summarize, in this thesis we have developed automated image analysis techniques for breast cancer assessment in DCE-MRI of the breast. We demonstrated that the developed methods perform well and have the potential to be incorporated in clinical practice to facilitate reading and interpretation of DCE-MRI.

Clinical applications

In this final section, we discuss the use in clinical practice of the algorithms developed in this thesis in order to facilitate the analysis of breast DCE-MRI and, consequently, improve the clinical outcome. Breast density assessment and automated detection of breast cancer, which have been extensively described in this thesis and will be further discussed in this section, are two clear clinical application examples. Other applications where the developed methods can serve as a tool to improve clinical workflow are automatic linkage of findings across current and prior examinations⁴³, multi-modal registration¹⁴¹, automated breast MRI quality assessment and pharmacokinetic modeling calibration⁵³, among others.

Automatic linkage of findings across current and prior examinations would allow clinical workstations to efficiently deal with prior MR images. Such a system would aid radiologists to locate suspicious areas in prior examinations. It has been shown that the inclusion of prior MR images in breast MRI reading reduces the rate of false positives associated with initial breast cancer MRI screening¹⁴². Similarly, registration among different image modalities would facilitate the interpretation and linkage of findings across modalities. To aid in these registration tasks, the correction of intra- and inter-patient signal intensity variability (bias field correction and image normalization) and the segmentation of different structures are essential to ensure high accuracy.

Another application is automated image quality assessment. To guarantee sufficient image quality, automated breast MRI quality controls can be incorporated during acquisition time. This assessment would guarantee that the quality of the image is sufficient to be interpreted by the radiologist. In breast MRI, motion artifacts and parenchyma or background enhancement are the main factors that create difficulties to interpret the images. Parenchyma or background enhancement refers to the enhancement of normal tissue due to pre-menopausal patients not being scanned during the optimal phase of the menstrual cycle¹⁶. The presence of these effects could alter the lesion kinetic curves, which are essential to provide an accurate diagnosis. Costs could be saved using a systematic quality assessment system. This is important as it is known that breast MRI is a relatively expensive imaging modality, especially compared to mammography or ultrasound. For instance, a breast MR scan that is clearly affected by motion artifacts could be repeated the same day. Furthermore, radiologists

would not need to spend time analyzing bad quality images that need to be repeated for a proper analysis. Minimum image quality criteria can be also defined as required by computer-aided detection systems. By doing so, the overall performance of a CAD system would increase as errors related to bad quality images would be avoided. The automated detection of these artifacts is a challenging task. The segmentation developed in this thesis could be used to segment body and breast structures in each time point of the DCE-MRI sequence. The segmentation of these structures is required for further processing and quantification of large patient movements or large variations of signal intensity along time.

A third clinical application where the developed segmentation methods can be used is pharmacokinetic modeling calibration. Pharmacokinetic modeling offers a quantitative approach to describe the temporal exchange of contrast agent in a tissue of interest. Measures such as tissue permeability and tissue vascularity, which can be used to characterize breast lesions¹⁴³, can be extracted using this approach. For a better interpretation of contrast enhancement lesions, researchers have tried to incorporate pharmacokinetic modeling to the interpretation of the DCE-MRI of the breast¹³⁹. Some of these models require calibrations with respect to reference tissues and make use of the signal intensity of specific region for determining physiological measures⁵³. In breast MRI, the pectoral muscle, which can be automatically segmented on each scan, can be used as a reference tissue given its properties. These pharmacokinetic modeling methods might have the ability to produce a robust quantification of contrast agent concentration where factors related to the patient MR scan, such as patient weight, administered dose and contrast medium used¹⁶, are taken into account.

Regarding the clinical applications described in this thesis, as mentioned earlier, breast density assessment is important as risk for developing breast cancer is four times larger in women with a breast density higher than 75%, compared to those with little or no density^{7,58}. In this thesis we used the developed breast density estimation algorithm to obtain reference standard measures in MRI and validate volumetric breast density obtained from mammography. This method could be also used for researchers who are interested in investigating the relationship of 3-dimensional fibroglandular tissue morphology and distribution patterns with risk for developing cancer¹⁴⁴.

Finally, computer-aided detection systems can be incorporated as a decision support tool into clinical workflow. For instance, a CAD system can be used as a prompting system to automatically marks suspicious areas during or after the first evaluation of the image. In this setting, the radiologist has to accept or neglect the suggested findings of the computer. Such a system can reduce the interpretation time of analyzing breast DCE-MRI data. Furthermore, as shown when CAD has been applied to other modalities⁹⁷⁻¹⁰⁰, otherwise missed lesions could be detected. In breast MRI, false-negative

diagnostic results are mainly categorized into overlooked and misinterpreted^{25,26,28}. Reported causes for observer errors are small lesion size, extensive background enhancement and malignant lesions with smooth margins, which is more commonly visible in benign disease. Although it has not been reported, it is evident that fatigue or lack of experience while analyzing 4-dimensional data are other causes for observer error.

A cancer detection system should provide a limited number of markers, otherwise radiologists will lose motivation to check them. In mammography and chest x-ray, commercial CAD systems have about 1-3 marks per case^{135,136} and 0.5-2 marks per image¹³⁷, respectively. While we are aware that an appropriate setting for breast DCE-MRI has to be determined using large multireader studies with a substantial amount of normal cases to represent the screening situation, we believe that 2 false positive per breast (4 false positives per study or volume) is acceptable as we are dealing with volumetric data that represents both breasts in a single volume. Considering the results achieved in this thesis, we are confident that the developed CAD system could be used as a prompting system in screening with breast MRI. Our CAD system achieved a sensitivity value for detection of malignant mass-like and non-mass-like lesions of 89% at 4 false-positive findings per normal patient.

Samenvatting

Borstkanker is wereldwijd de meest voorkomende doodsoorzaak door kanker bij vrouwen¹. Het vroegtijdig opsporen van borstkanker verhoogt de overlevingskans. Veel landen hebben daarom een borstkankerscreening programma ingevoerd. In reguliere programma's voor borstkankerscreening worden asymptotische vrouwen eens in de paar jaar uitgenodigd voor het maken van een mammogram. Echter, mammografie kent beperkingen. De aanwezigheid van borstweefsel met een hoge densiteit kan de tumor op de röntgenfoto namelijk vertroebelen^{10,58,76}. Verder is er geconstateerd dat een hoge borstdensiteit een risicofactor is voor de ontwikkeling van borstkanker¹⁰. Andere risicofactoren zijn het voorkomen van borstkanker in de familie en de aanwezigheid van een kiembaanmutatie van het *BRCA1* of *BRCA2* gen^{13,14}.

Een benadering om deze tekortkomingen te ondervangen en de borstkankerscreening te optimaliseren is de invoering van een gepersonaliseerde screening¹¹. In gepersonaliseerde borstkankerscreening programma's worden vrouwen met een hoog risico op het ontwikkelen van borstkanker gescreend met een complementaire modaliteit. Dynamic Contrast-Enhanced Magnetic Resonance (DCE-MRI) heeft potentie laten zien als screeningsinstrument voor de populatie vrouwen met een hoog risico op het ontwikkelen van borstkanker¹⁷⁻²². In vergelijking met mammografie heeft DCE-MRI een hogere sensitiviteit, vooral bij vrouwen met een hoge borstweefsel densiteit²³. MRI is bovendien geschikt voor het opsporen van hoog gradige ductale carcinoom in situ (DCIS), waarvan bekend is dat het zich sneller ontwikkelt tot een invasieve kanker dan laag gradige DCIS¹¹⁸. Echter, het analyseren van de data is een tijdrovende taak. De analyse van borst MRI data vereist namelijk een interpretatie van vier-dimensionaal DCE-data, evenals de correlatie tussen multi-parametrische data van andere MRI sequenties. Bovendien rapporteren recente studies dat er zichtbare kankers in een voorgaand MRI onderzoek werden gevonden die eerder niet gezien waren of fout waren geïnterpreteerd^{25,26}.

Deze thesis focust zich op de automatische analyse van borst MRI. Automatische beeldanalyse technieken en detectie systemen met behulp van een computer zijn nodig om de klinische werkwijze te verbeteren en de radioloog te assisteren bij het lezen en interpreteren van MRI beelden.

Hoofdstuk 2 beschrijft de ontwikkeling van twee volledig geautomatiseerde methoden voor de segmentatie van de pectoralis spier in borst MRI. De identificatie van de pectoralis spier is een belangrijke stap in methoden voor automatische borstkanker analyse. Er zijn twee mogelijke toepassingen. De eerste is het vinden van de grens tussen het lichaam en de borst, waarbij de meest uitdagende taak het segmenteren van de borst is door de grote verschillen tussen patiënten. De tweede methode is het gebruik van de pectoralis spier als referentie weefsel voor farmacokinetische modellen. De voorgestelde methoden voor de segmentatie van de pectoralis houden rekening met

de 3-dimensionale beeldvorming van borst MRI en zijn gebaseerd op een atlas-based methode: een probabilistische en een multi-atlas methode. Beide methoden zijn geëvalueerd met dezelfde dataset. De multi-atlas methode presteert iets beter dan de probabilistische (respectievelijk gemiddelde DSC waarden van 0.74 ± 0.06 en 0.72 ± 0.09) maar de computerberekeningen duurden 14 keer langer.

Hoofdstuk 3 beschrijft een automatische methode om de borst te segmenteren in borst MRI en de borstdensiteit te schatten. Andere beeld verwerkingsmethoden om deze taken uit te voeren worden ook in dit hoofdstuk beschreven, zoals beeld normalisatie, bias-field correctie en sternum detectie. De borst wordt gesegmenteerd door het automatisch identificeren van de grens tussen borst en lichaam en tussen borst en achtergrond. De segmentatie van de borst en lichaam wordt uitgevoerd door gebruik te maken van de probabilistische atlas benadering, zoals beschreven in hoofdstuk 2. Twee-dimensionale region growing wordt toegepast om de borst van de achtergrond te segmenteren. Vervolgens wordt het borst volume van elke borst opnieuw gedefinieerd door gebruik te maken van de detectie van de borstcurvatuur en herkenningspunten. Zo worden delen met vetweefsel, die geen onderdeel vormen van de borst en niet te zien zijn in mammogrammen, uitgesloten. Hierdoor zijn de metingen van de borstdensiteit berekend met MRI vergelijkbaar met borstdensiteit berekend met mammografie. Tenslotte is de borstdensiteit voor elke borst onafhankelijk berekend door gebruik te maken van Expectation-Maximization algoritme en skin-fold removal. Resultaten laten een grote overeenkomst zien tussen automatische borst segmentatie en borstdensiteit en handmatige segmentatie door experts. Voor borst segmentatie, bereikten we een gemiddelde DSC waarde 0.94 ± 0.04 en een gemiddelde totale overlap van 0.96 ± 0.02 . Voor de automatische fibroglandular weefsel segmentatie, bereikten we een gemiddelde DSC waarde van 0.80 ± 0.13 .

Hoofdstuk 4 beschrijft een licht bewerkte methode ten opzichte van de methode met de borst segmentatie en berekening van de borstdensiteit beschreven in hoofdstuk 3. Deze methode wordt gebruikt om Volpara 1.4.3 (Mātakina, Wellington, New Zealand) te valideren, wat een commercieel beschikbare methode is om volumetrische borstdensiteit te meten in FFDM. Volumetrische berekeningen verkregen met Volpara door middel van FFDM beelden werden vergeleken met volume berekeningen verkregen met borst MRI data van dezelfde patiënt. Voor deze vergelijking werd de door ons ontwikkelde methode toegepast op 250 MRI-scans. Een radioloog met expertise in borst imaging beoordeelde nauwkeurig alle slabs van de borst segmentatie en selecteerde 186 (74,4%) MRI segmentatie studies als geschikt om als referentie te gebruiken voor de validatie van de FFDM densiteit metingen. Volumetrische borstdensiteit en volume van borstweefsel verkregen met Volpara laten hoge correlaties zien wanneer deze worden vergeleken met MRI metingen (correlatie waarden per borst en per studie van 0.91

en 0.93, respectievelijk). Deze bevinding laat zien dat het haalbaar is om accurate metingen te verrichten van absolute en relatieve volumes van borstweefsel van FFDM densiteit metingen. De beschikbaarheid van deze metingen is cruciaal voor de ontwikkeling van objectieve borstkanker modellen en kan, in plaats van de stratificatie van vrouwen, gebruikt worden voor gepersonaliseerde borstkankerscreening protocollen.

Hoofdstuk 5 geeft een beschrijving van een Computer-Aided Detection (CAD) systeem voor het automatisch opsporen van borstkanker in DCE-MRI. Het CAD systeem corrigeert fouten veroorzaakt door beweging en segmenteert automatisch de borst. Vervolgens worden laesie locatie kandidaten opgespoord door gebruik te maken van relatieve contrast en weefsel structuur eigenschappen zoals blob descriptors. De laatste classificatie wordt gemaakt door gebruik te maken van region-based morfologische en kinetische eigenschappen berekend met gesegmenteerde laesie kandidaten. De kwaliteit van het systeem voor de opsporing van massa- en nietmassa-achtige laesies is geëvalueerd. Borstkanker wordt gedetecteerd met een hoge sensitiviteit bij een acceptabel aantal fout-positieven: 0.55, 0.78, 0.89 en 0.95 bij 0.1, 1, 4 en 7 fout positieven per scan, waarbij fout-positieven werden bepaald bij normale patiënten. Ook bleek dat correctie voor beweging nodig is om de prestatie van het CAD-systeem te optimaliseren. De verkregen resultaten zijn veelbelovend. Het systeem kan gebruikt worden als hulpmiddel voor radiologen tijdens de analyse van DCE-MRI data door het automatisch tonen van verdachte gebieden.

In hoofdstuk 6 onderzochten we of ons CAD systeem borstkanker kan opsporen in de gevallen waar de radiologen bij borstkanker screening met MRI het niet hebben ontdekt of verkeerd geïnterpreteerd hebben. Er werd een dataset van 41 paren van DCE-MRI studies verzameld. Elk paar bestond uit een voorgaande DCE-MRI en een follow-up DCE-MRI waarbij borstkanker werd gediagnosticeerd. In een retrospectieve review van deze casussen vonden wij dat in 16 gevallen de borstkanker zichtbaar was en 15 gevallen waar het minimaal zichtbaar was in de eerdere DCE-MRI scan. Ons CAD-systeem ontdekte 63% van de zichtbare en 19% van de minimaal zichtbare laesies bij 2 fout positieven per borst. Deze resultaten suggereren dat de integratie van dit systeem in de klinische praktijk de radioloog kan assisteren om fouten bij het lezen te voorkomen.

Samenvattend, in deze thesis hebben we automatische beeld analyse technieken voor de opsporing van borstkanker ontwikkeld in DCE-MRI van de borst. We hebben laten zien dat de ontwikkelde methoden goed presteren en de potentie hebben om geïmplementeerd te worden in de klinische praktijk om het lezen en interpreteren van DCE-MRI te faciliteren.

Resum

El càncer de mama, a nivell mundial, és la causa de mort més comuna entre les dones¹. Donat que la detecció precoç d'aquesta malaltia augmenta les possibilitats de supervivència, s'han introduït en molts països programes de cribratge de càncer de mama (o pit). En aquests programes de cribratge s'examinen de forma periòdica dones asimptomàtiques mitjançant mamografia digital. No obstant, la mamografia té les seves limitacions i és que els tumors poden quedar ocults per la presència de teixit dens^{10,58,76}. Aquest fet dificulta la detecció de càncer en determinats grups de pacients, especialment en aquelles dones amb mames denses. La presència de teixit dens, o altrament anomenat densitat mamària, també ha estat reconegut com un factor de risc per desenvolupar càncer de mama¹⁰. Altres factors que s'associen amb un major risc són els antecedents familiars en aquesta malaltia i la presència de mutacions germinals en els gens *BRCA1* o *BRCA2*^{13,14}.

Per tal de superar aquestes deficiències i optimitzar la detecció del càncer de mama, s'estan introduït programes de cribratge personalitzats¹¹. En aquests programes personalitzats, s'examinen les dones amb alt risc de desenvolupar càncer de mama utilitzant modalitats d'imatge complementàries a la mamografia. La imatge de ressonància magnètica, MRI (*Magnetic Resonance Imaging*) de l'anglès, amb seqüència dinàmica amb agent contrast, DCE-MRI (*Dynamic Contrast-Enhanced Magnetic Resonance Imaging*) de l'anglès, és una eina molt eficient per a la detecció de càncer de mama en poblacions d'alt risc. En comparació amb la mamografia, DCE-MRI té una major sensibilitat¹⁷⁻²², especialment en dones amb mames denses²³. D'altra banda, DCE-MRI és particularment útil per a la detecció de carcinomes ductals in situ, DCIS (*Ductal Carcinoma In Situ*) de l'anglès, de grau elevat, els quals progressen més ràpidament cap a una malaltia invasiva que els DCIS de grau baix¹¹⁸. Malauradament, l'anàlisi d'imatges de DCE-MRI de la mama és una tasca feixuga perquè requereix la interpretació de dades dinàmiques en quatre dimensions, així com la correlació de dades multi-paramètriques que provenen d'altres seqüències de MRI. A més a més, estudis recents han demostrat l'existència d'errors per part dels observadors experts a causa d'una pobra interpretació de lesions de càncer de mama en DCE-MRI^{25,26}.

Per optimitzar i facilitar la lectura i interpretació d'imatges DCE-MRI als radiòlegs especialitzats calen sistemes d'ajuda assistits per ordinador per la detecció i avaluació del càncer de mama. Amb tal finalitat, en aquesta tesi s'han estudiat i desenvolupat noves tècniques informàtiques per l'anàlisi automatitzat d'imatges de DCE-MRI.

La identificació i la delineació de les diferents estructures visibles en la imatge, el múscul pectoral entre elles, és essencial pels sistemes d'ajuda assistits per ordinador per la detecció del càncer de mama. El capítol 2 es centra en el desenvolupament de dos mètodes totalment automatitzats per a la segmentació del múscul pectoral en DCE-MRI. La segmentació del múscul pectoral és útil per obtenir la delimitació de la superfície

definida entre el cos i els pits. A més a més, el múscul pectoral pot ésser utilitzat com a teixit de referència en la modelització de paràmetres farmacocinètics. Aquests poden aportar informació rellevant per a la classificació de lesions en lesions malignes o lesions benignes. Els mètodes de segmentació proposats en aquesta tesi tenen en compte la naturalesa en 3 dimensions de les imatges de MRI i utilitzen informació provinent d'atles: un mètode d'atles probabilístic i un mètode multi-atlas. Ambdós mètodes han estat avaluats en el mateix conjunt de dades. La metodologia multi-atlas va superar lleugerament el mètode probabilístic amb valors mitjans de *Dice Similarity Coefficient* (DSC) de 0.74 ± 0.06 i 0.72 ± 0.09 , respectivament, però el temps computacional va ser 14 vegades més gran.

En el capítol 3 es presenta un mètode automatitzat per segmentar la mama i estimar la quantitat de teixit dens de la mateixa. Per dur a terme aquesta tasca, s'han desenvolupat i descrit en aquest capítol altres tècniques de processament d'imatge necessaris, com la normalització i la correcció de la intensitat de la imatge, i algorismes per a la detecció de l'estèrnum. El pit es segmenta mitjançant la identificació automàtica de la superfície del múscul pectoral i la superfície de la mama. La segmentació de la superfície del múscul pectoral es realitza utilitzant la tècnica d'atles probabilístic explicada en el capítol 2. L'algorisme *region growing* s'aplica per definir la superfície del pit. Posteriorment, per tal d'excloure àrees de teixit gras que no formen part de la mama i que no són visibles en imatges de mamografia, es redefineix el volum de cada pit emprant informació de curvatura i la posició de l'estèrnum. D'aquesta manera, les mesures de densitat extrems de MRI són comparables a les estimacions de densitat de la mama obtingudes en mamografia. Finalment, es calcula la densitat mamària de forma independent per a cada pit utilitzant l'algorisme *Expectation-Maximization* i un algorisme d'eliminació d'artefactes produïts per plecs en la pell durant el posicionament de la dona en l'escàner. Els resultats obtinguts mostren alta concordança entre les segmentacions del pit i del teixit dens produïdes per l'algorisme proposat i les anotacions manual realitzades per experts. Per a la segmentació de la mama, es va obtenir un valor DSC de mitjana de 0.94 ± 0.04 i un valor de *total overlap* de 0.96 ± 0.02 . Per a la segmentació automàtica del teixit dens, es va obtenir un valor DSC de mitjana de 0.80 ± 0.13 .

Una versió lleugerament modificada d'aquest algorisme de segmentació de pit i estimació de densitat mamària s'ha utilitzat en l'estudi descrit en el capítol 4. Aquest estudi consisteix en la validació de Volpara 1.4.3 (Mātakina, Wellington, New Zealand), que és un producte informàtic comercial per calcular mesures volumètriques de densitat mamària en mamografia digital. Les estimacions volumètriques obtingudes en imatges de mamografia utilitzant Volpara es van comparar amb les mesures volumètriques obtingudes a partir de dades de MRI de la mama del mateix pacient. Per a aquest propòsit,

es va aplicar el mètode desenvolupat en aquesta tesi a un conjunt de dades compost de 250 imatges de MRI. Un radiòleg amb experiència en càncer de pit va revisar de forma acurada totes les segmentacions i va notificar que les segmentacions obtingudes en 186 (74.4%) casos eren apropiades per calcular mesures de densitat mamària de referència. Es va observar que els valors volumètrics de densitat mamària calculats amb Volpara presenten una alta correlació en comparació amb les mesures obtingudes en MRI (valors de correlació per pit i per imatge de 0.91 i 0.93, respectivament). Aquesta troballa demostra que és factible estimar de forma precisa la densitat mamària en mamografia digital. La disponibilitat d'aquest tipus de mesures és essencial per al desenvolupament de models de risc de càncer de mama i poden ésser utilitzades per a l'estratificació de les dones en els programes de monitorització personalitzats.

En el capítol 5 es descriu un sistema d'ajuda de detecció assistit per ordinador, CAD (*Computer-Aided Detection*) de l'anglès, per detectar de forma automàtica càncer de mama en imatges de DCE-MRI. El sistema CAD comença corregint els artefactes produïts pel moviment del pacient durant l'adquisició de la imatge i a continuació segmenta els pits. Posteriorment, es detecten els candidats a pertànyer a una lesió utilitzant característiques de realç relatiu d'intensitats i característiques de textura que inclouen descriptors de *blobs*. La classificació final de cada regió candidata a lesió es duu a terme emprant característiques morfològiques i característiques cinètiques. Amb els experiments proposats s'estudia el rendiment del sistema CAD per a la detecció de tot tipus de lesions de càncer de mama. Com a resultat, el sistema CAD detecta càncer de mama a una alta sensibilitat amb un nombre acceptable de falsos positius: 0.55, 0.78, 0.89 i 0.95 a 0.1, 1, 4 i 7 falsos positius per pacient normal. També s'observa que la correcció dels artefactes produïts pel moviment del pacient durant l'adquisició és important per aconseguir el màxim rendiment del sistema CAD. En general, els resultats obtinguts són prometedors i indicatius que el sistema CAD proposat pot ésser utilitzat com a eina de suport durant l'anàlisi de dades de DCE-MRI.

Finalment, en el capítol 6, s'estudia el potencial del sistema CAD desenvolupat en aquesta tesi per detectar lesions de càncer de mama que es van passar per alt o es van malinterpretar en un programa de cribratge amb MRI. Per a la realització d'aquest estudi, es van obtenir un total de 41 parells d'imatges DCE-MRI. Cada parell estava compost per la imatge DCE-MRI on el càncer va ser detectat (d'aquí en endavant DCE-MRI de diagnòstic) i una imatge de DCE-MRI del mateix pacient anterior a la DCE-MRI de diagnòstic (DCE-MRI prèvia al DCE-MRI de diagnòstic) on el càncer no va ser detectat. En una revisió de forma retrospectiva de totes les imatges de DCE-MRI, es va trobar que 16 càncers de mama eren visibles i 15 eren mínimament visibles en les imatges de DCE-MRI prèvies a les DCE-MRI de diagnòstic. El sistema CAD va ser capaç de detectar 63% i 19% de les lesions que eren visibles i mínimament visibles

de forma retrospectiva, respectivament, a 2 falsos positius per pit. Aquests resultats suggereixen que la integració d'un sistema CAD d'aquest tipus en la pràctica clínica pot ajudar els radiòlegs per evitar errors de lectura en la detecció de càncer de mama en DCE-MRI.

En resum, en aquesta tesi s'han desenvolupat tècniques automatitzades d'anàlisi d'imatge per a l'avaluació i la detecció de càncer de mama en imatges de DCE-MRI. S'ha demostrat que els mètodes proposats ofereixen un bon rendiment. La seva integració en la pràctica clínica podria facilitar i optimitzar la lectura i la interpretació d'imatges DCE-MRI.

Publications

Papers in international journals

J. Schwaab, Y. Diez, A. Oliver, S. Durand, R. Martí, J. van Zelst, **A. Gubern-Mérida**, A. B. Mourri, J. Gregori and M. Günther. “Automated quality assessment in 3D breast ultrasound images”, *submitted*.

S. M. M. R. Al-Arif, **A. Gubern-Mérida** and R. Martí. “Evaluation of bias field correction methods on breast MRI”, *submitted*.

T. Tan, J.J. Mordang, J. van Zelst, A. Grivegnée, **A. Gubern-Mérida**, J. Melendez, R. M. Mann, W. Zhang, B. Platel and N. Karssemeijer. “Computer-aided detection of breast cancers using Haar-like features in automated 3D breast ultrasound”, *Medical Physics*, to appear.

A. Gubern-Mérida, R. Martí, J. Melendez, J. L. Hauth, R. M. Mann, N. Karssemeijer and B. Platel. “Automated localization of breast cancer in DCE-MRI”, *Medical Image Analysis*, 20(1): 265–274, 2015.

A. Gubern-Mérida, M. Kallenberg, R. M. Mann, R. Martí and N. Karssemeijer. “Breast segmentation and density estimation in breast MRI: a fully automatic framework”, *IEEE Journal of Biomedical and Health Informatics*, 19(1): 349–357, 2015.

A. Gubern-Mérida, M. Kallenberg, B. Platel, R. M. Mann, R. Martí and N. Karssemeijer. “Volumetric breast density estimation from Full-Field Digital Mammograms: a validation study”, *PLoS One*, 9(1): e85952, 2014.

Papers in conference proceedings

L. Wang, J. Strehlow, J. Ruehaak, F. Weiler, Y. Diez, **A. Gubern-Mérida**, S. Diekmann, H. Laue and H. K. Hahn. “A fast alignment method for breast MRI follow-up studies using automated breast segmentation and current-prior registration”, *Medical Imaging of Proceedings of the SPIE*, 2015.

Y. Diez, **A. Gubern-Mérida**, L. Wang, S. Diekmann, J. Martí, B. Platel, J. Kramme and R. Martí. “Comparison of Methods for Current-to-Prior Registration of Breast DCE-MRI”, *IWDM 2014, volume 8539 of Proceedings of the 12th international workshop on Digital Mammography*, pages 689–695, 2014.

A. Gubern-Mérida, B. Platel, R. M. Mann, R. Martí and N. Karssemeijer. “Automated localization of malignant lesions in breast DCE-MRI”, *MICCAI Workshop: Breast Image Analysis*, 2013.

A. Gubern-Mérida, L. Wang, M. Kallenberg, R. Martí, H.K. Hahn and N. Karssemeijer. “Breast segmentation in MRI: quantitative evaluation of three methods”, *Medical Imaging, volume 8669 of Proceedings of the SPIE*, pages 86693G-86693G-7, 2013.

A. Gubern-Mérida, M. Kallenberg, R. Martí and N. Karssemeijer. “Segmentation of the pectoral muscle in breast MRI using atlas-based approaches”, *Medical Imaging Computing and Computer-Assisted Intervention - MICCAI 2012*, volume 7511 of Lecture Notes in Computer Science, pages 371-378, 2012.

A. Gubern-Mérida, M. Kallenberg, R. Martí and N. Karssemeijer. “Fully automatic fibroglandular tissue segmentation in breast MRI: atlas-based approach”, *MICCAI Workshop: Breast Image Analysis*, 2011.

A. Gubern-Mérida, M. Kallenberg, R. Martí and N. Karssemeijer. “Multi-class probabilistic atlas-based segmentation method in breast MRI”, *5th Iberian Conference on Pattern Recognition and Image Analysis, IbPRIA 2011*, volume 6669 of Lecture Notes in Computer Science, pages 660-667, 2011.

Abstracts in conference proceedings

A. Gubern-Mérida, S. Vreemann, R. Martí, J. Melendez, S. Lardenoiije, R. M. Mann, B. Platel and N. Karssemeijer. “Automated detection of breast cancer as an aid in the interpretation of screening MRI”, *European Congress of Radiology*, 2015.

S. Vreemann, **A. Gubern-Mérida**, S. Lardenoiije, B. Platel, N. Karssemeijer and R. M. Mann. “A critical audit of a breast MRI screening programme for intermediate and high risk patients in clinical practice”, *European Congress of Radiology*, 2015.

L. Wang, **A. Gubern-Mérida**, O. Diaz, Y. Diez, R. M. Mann, S. Diekmann, F. Zöhrer, H. Laue and J. Schwaab. “Automated detection of motion in breast DCE-MRI to assess study quality and prevent unnecessary call-backs”, *European Congress of Radiology*, 2015.

M. U. Dalmis, **A. Gubern-Mérida**, S. Vreemann, B. Platel, R. M. Mann and N. Karssemeijer. “Is late-phase information necessary for dynamic evaluation of breast cancer?”, *European Congress of Radiology*, 2015.

Bibliography

- [1] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics," *CA Cancer J Clin*, vol. 61, pp. 69–90, 2011.
- [2] J. Ferlay, E. Steliarova-Foucher, J. Lortet-Tieulent, S. Rosso, J. W. W. Coebergh, H. Comber, D. Forman, and F. Bray, "Cancer incidence and mortality patterns in europe: estimates for 40 countries in 2012.," *Eur J Cancer*, vol. 49, pp. 1374–1403, Apr 2013.
- [3] D. Schopper and C. de Wolf, "How effective are breast cancer screening programmes by mammography? Review of the current evidence," *Eur J Cancer*, vol. 45, pp. 1916–1923, 2009.
- [4] G. van Schoor, *Breast cancer screening: Increasing benefits and extending scope*. PhD thesis, Radboud University Nijmegen, 2012.
- [5] M. T. Mandelson, N. Oestreicher, P. L. Porter, D. White, C. A. Finder, S. H. Taplin, and E. White, "Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers," *J Natl Cancer Inst*, vol. 92, pp. 1081–1087, 2000.
- [6] J. J. Heine and P. Malhotra, "Mammographic tissue, breast cancer risk, serial image analysis, and digital mammography. part 1. tissue and related risk factors," *Acad Radiol*, vol. 9, pp. 298–316, 2002.
- [7] N. F. Boyd, H. Guo, L. J. Martin, L. Sun, J. Stone, E. Fishell, R. A. Jong, G. Hislop, A. Chiarelli, S. Minkin, and M. J. Yaffe, "Mammographic density and the risk and detection of breast cancer," *N Engl J Med*, vol. 356, pp. 227–236, 2007.
- [8] J. Nederend, L. E. Duijm, A. C. Voogd, J. H. Groenewoud, F. H. Jansen, and M. W. Louwman, "Trends in incidence and detection of advanced breast cancer at biennial screening mammography in the Netherlands: a population based study," *Breast Cancer Res*, vol. 14, p. R10, 2012.
- [9] K. Kerlikowske, "The mammogram that cried wolfe," *N Engl J Med*, vol. 356, pp. 297–300, 2007.
- [10] N. F. Boyd, L. J. Martin, M. J. Yaffe, and S. Minkin, "Mammographic density and breast cancer risk: current understanding and future prospects," *Breast Cancer Res*, vol. 13, p. 223, 2011.
- [11] J. T. Schousboe, K. Kerlikowske, A. Loh, and S. R. Cummings, "Personalizing mammography by breast density and other risk factors for breast cancer: analysis

- of health benefits and cost-effectiveness,” *Ann Intern Med*, vol. 155, pp. 10–20, 2011.
- [12] W. A. Berg, “Tailored supplemental screening for breast cancer: What now and what next?,” *AJR Am J Roentgenol*, vol. 192, pp. 390–399, 2009.
- [13] A. J. Rijnsburger, I.-M. Obdeijn, R. Kaas, M. M. Tilanus-Linthorst, C. Boetes, C. E. Loo, M. N. Wasser, E. Bergers, T. Kok, S. H. Muller, H. Peterse, R. A. Tollenaar, N. Hoogerbrugge, S. Meijer, C. C. Bartels, C. Seynaeve, M. J. Hooning, M. Kriege, P. I. M. Schmitz, J. C. Oosterwijk, H. J. de Koning, E. J. Rutgers, and J. G. Klijn, “BRCA1-associated breast cancers present differently from BRCA2-associated and familial cases: Long-term follow-up of the Dutch MRISC screening study,” *J Clin Oncol*, vol. 28, pp. 5265–5273, 2010.
- [14] J. Klijn, “Early diagnosis of hereditary breast cancer by magnetic resonance imaging: What is realistic?,” *J Clin Oncol*, vol. 28, pp. 1441–1445, 2010.
- [15] D. Saslow, C. Boetes, W. Burke, S. Harms, M. O. Leach, C. D. Lehman, E. Morris, E. Pisano, M. Schnall, S. Sener, R. A. Smith, E. Warner, M. Yaffe, K. S. Andrews, C. A. Russell, and American Cancer Society Breast Cancer Advisory Group, “American cancer society guidelines for breast screening with MRI as an adjunct to mammography,” *CA Cancer J Clin*, vol. 57, pp. 75–89, 2007.
- [16] R. M. Mann, C. K. Kuhl, K. Kinkel, and C. Boetes, “Breast MRI: Guidelines from the European Society of Breast Imaging,” *Eur Radiol*, vol. 18, pp. 1307–1318, 2008.
- [17] M. Kriege, C. T. Brekelmans, C. Boetes, P. E. Besnard, H. M. Zonderland, I. M. Obdeijn, R. A. Manoliu, T. Kok, H. Peterse, M. M. Tilanus-Linthorst, S. H. Muller, S. Meijer, J. C. Oosterwijk, L. V. Beex, R. A. Tollenaar, H. J. de Koning, E. J. Rutgers, and J. G. Klijn, “Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition,” *New England Journal of Medicine*, vol. 351, no. 5, pp. 427–437, 2004.
- [18] C. D. Lehman, J. D. Blume, P. Weatherall, D. Thickman, N. Hylton, E. Warner, E. Pisano, S. J. Schnitt, C. Gatsonis, M. Schnall, and for the International Breast MRI Consortium Working Group, “Screening women at high risk for breast cancer with mammography and magnetic resonance imaging,” *Cancer*, vol. 103, no. 9, pp. 1898–1905, 2005.
- [19] E. Warner, D. B. Plewes, K. A. Hill, P. A. Causer, J. T. Zubovits, R. A. Jong, M. R. Cutrara, G. DeBoer, M. J. Yaffe, S. J. Messner, W. S. Meschino, C. A.

- Piron, S. A. Narod, "Surveillance of BRCA1 and BRCA2 mutation carriers with magnetic resonance imaging, ultrasound, mammography, and clinical breast examination," *JAMA*, vol. 292, no. 11, pp. 1317–1325, 2004.
- [20] E. A. Morris, L. Liberman, D. J. Ballon, M. Robson, A. F. Abramson, A. Heerdt, and D. D. Dershaw, "MRI of occult breast carcinoma in a high-risk population.," *AJR Am J Roentgenol*, vol. 181, no. 3, pp. 619–26, 2003.
- [21] F. J. Gilbert, R. M. L. Warren, G. Kwan-Lim, D. J. Thompson, R. A. Eeles, D. G. Evans, and M. O. a. Leach, "Cancers in BRCA1 and BRCA2 carriers and in women at high risk for breast cancer: MR imaging and mammographic features," *Radiology*, vol. 252, no. 2, pp. 358–368, 2009.
- [22] R. M. Mann, Y. L. Hoogeveen, J. G. Blickman, and C. Boetes, "MRI compared to conventional diagnostic work-up in the detection and evaluation of invasive lobular carcinoma of the breast: A review of existing literature," *Breast Cancer Res Treat*, vol. 107, pp. 1–14, 2008.
- [23] E. D. Pisano, R. E. Hendrick, M. J. Yaffe, J. K. Baum, S. Acharyya, J. B. Cormack, L. A. Hanna, E. F. Conant, L. L. Fajardo, L. W. Bassett, C. J. D'Orsi, R. A. Jong, M. Rebner, A. N. A. Tosteson, C. A. Gatsonis, and for the DMIST Investigators Group, "Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in DMIST," *Radiology*, vol. 246, pp. 376–383, 2008.
- [24] N. H. Peters, I. H. Rinkes, N. P. Zuithoff, W. P. Mali, K. G. Moons, and P. H. Peeters, "Meta-analysis of MR imaging in the diagnosis of breast lesions," *Radiology*, vol. 246, pp. 116–124, 2008.
- [25] E. B. Pages, I. Millet, D. Hoa, F. C. Doyon, and P. Taourel, "Undiagnosed breast cancer at MR imaging: analysis of causes," *Radiology*, vol. 264, pp. 40–50, Jul 2012.
- [26] K. Yamaguchi, D. Schacht, G. M. Newstead, A. R. Bradbury, M. S. Verp, O. I. Olopade, and H. Abe, "Breast cancer detected on an incident (second or subsequent) round of screening MRI: MRI features of false-negative cases," *AJR Am J Roentgenol*, vol. 201, pp. 1155–1163, 2013.
- [27] C. K. Kuhl, P. Mielcareck, S. Klaschik, C. Leutner, E. Wardelmann, J. Gieseke, and H. H. Schild, "Dynamic breast MR imaging: Are signal intensity time course data useful for differential diagnosis of enhancing lesions?," *Radiology*, vol. 211, no. 1, pp. 101–110, 1999.

-
- [28] I.-M. Obdeijn, C. E. Loo, A. J. Rijnsburger, M. N. Wasser, E. Bergers, T. Kok, J. G. Klijn, and C. Boetes, "Assessment of false-negative cases of breast MR imaging in women with a familial or genetic predisposition.," *Breast Cancer Research and Treatment*, vol. 119, pp. 399–407, Jan 2010.
- [29] E. Morris, C. Comstock, and C. Lee, "ACR BI-RADS Magnetic Resonance Imaging," in *2013 ACR BI-RADS Atlas: Breast Imaging Reporting and Data System*, American College of Radiology, 2014.
- [30] M. D. Dorrius, M. C. Jansen-van der Weide, P. M. A. van Ooijen, R. M. Pijnappel, and M. Oudkerk, "Computer-aided detection in breast MRI: A systematic review and meta-analysis," *Eur Radiol*, vol. 21, pp. 1600–1608, 2011.
- [31] D. Newell, K. Nie, J.-H. Chen, C.-C. Hsu, H. J. Yu, O. Nalcioglu, and M.-Y. Su, "Selection of diagnostic features on breast MRI to differentiate between malignant and benign lesions using computer-aided diagnosis: differences in lesions presenting as mass and non-mass-like enhancement," *Eur Radiol*, vol. 20, pp. 771–781, 2010.
- [32] S. M. Astley, "Computer-aided detection for screening mammography," *Acad Radiol*, vol. 11, pp. 1139–1143, 2004.
- [33] T. Onega, E. J. A. Bowles, D. L. Miglioretti, P. A. Carney, B. M. Geller, B. C. Yankaskas, K. Kerlikowske, E. A. Sickles, and J. G. Elmore, "Radiologists' perceptions of computer aided detection versus double reading for mammography interpretation," *Acad Radiol*, vol. 17, pp. 1217–1226, 2010.
- [34] R. F. Brem, J. Baum, M. Lechner, S. Kaplan, S. Souders, L. G. Naul, and J. Hoffmeister, "Improvement in sensitivity of screening mammography with computer-aided detection: a multiinstitutional trial," *AJR Am J Roentgenol*, vol. 181, pp. 687–693, 2003.
- [35] F. J. Gilbert, S. M. Astley, M. G. C. Gillan, O. F. Agbaje, M. G. Wallis, J. James, C. R. M. Boggis, S. W. Duffy, and CADET II Group, "Single reading with computer-aided detection for screening mammography," *N Engl J Med*, vol. 359, pp. 1675–1684, 2008.
- [36] M. J. Morton, D. H. Whaley, K. R. Brandt, and K. K. Amrami, "Screening mammograms: interpretation with computer-aided detection—prospective evaluation," *Radiology*, vol. 239, pp. 375–383, 2006.

- [37] M. Samulski, R. Hupse, C. Boetes, R. Mus, G. den Heeten, and N. Karssemeijer, "Using Computer Aided Detection in Mammography as a Decision Support," *Eur Radiol*, vol. 20, pp. 2323–2330, 2010.
- [38] A. Vignati, V. Giannini, M. De Luca, L. Morra, D. Persano, L. A. Carbonaro, I. Bertotto, L. Martincich, D. Regge, A. Bert, and F. Sardanelli, "Performance of a fully automatic lesion detection system for breast DCE-MRI," *J Magn Reson Imaging*, vol. 34, pp. 1341–1351, 2011.
- [39] D. M. Renz, J. Böttcher, F. Diekmann, A. Poellinger, M. H. Maurer, A. Pfeil, F. Streitparth, F. Collettini, U. Bick, B. Hamm, and E. M. Fallenberg, "Detection and classification of contrast-enhancing masses by a fully automatic computer-assisted diagnosis system for breast MRI," *J Magn Reson Imaging*, vol. 35, pp. 1077–1088, 2012.
- [40] Y.-C. Chang, Y.-H. Huang, C.-S. Huang, J.-H. Chen, and R.-F. Chang, "Computerized breast lesions detection using kinetic and morphologic analysis for dynamic contrast-enhanced MRI," *J Magn Reson Imaging*, vol. 32, pp. 514–522, 2014.
- [41] L. Wang, B. Platel, T. Ivanovskaya, M. Harz, and H. Hahn, "Fully automatic breast segmentation in 3D breast MRI," in *IEEE Int Symp Biomedical Imaging*, 2012.
- [42] C. Gallego Ortiz and A. L. Martel, "Automatic atlas-based segmentation of the breast in MRI for 3D breast volume computation," *Med Phys*, vol. 39, pp. 5835–5848, 2012.
- [43] Y. Diez, A. Gubern-Mérida, L. Wang, S. Diekmann, J. Martí, B. Platel, J. Kramme, and R. Martí, "Comparison of methods for current-to-prior registration of breast DCE-MRI," in *IWDM '14: Proceedings of the 12th international workshop on Digital Mammography*, vol. 8539, (Berlin, Heidelberg), pp. 689–695, Springer-Verlag, 2014.
- [44] K. Nie, J.-H. Chen, S. Chan, M.-K. I. Chau, H. J. Yu, S. Bahri, T. Tseng, O. Nalcioglu, and M.-Y. Su, "Development of a quantitative method for analysis of breast density based on three-dimensional breast MRI," *Med Phys*, vol. 35, pp. 5253–5262, 2008.
- [45] S. van Engeland, P. R. Snoeren, H. Huisman, C. Boetes, and N. Karssemeijer, "Volumetric breast density estimation from full-field digital mammograms," *IEEE Trans Med Imaging*, vol. 25, pp. 273–282, 2006.

-
- [46] R. Highnam, M. Brady, M. J. Yaffe, N. Karssemeijer, and J. Harvey, "Robust breast composition measurement - Volpara," in *IWDM '10: Proceedings of the 10th international workshop on Digital Mammography*, (Berlin, Heidelberg), pp. 342–349, Springer-Verlag, 2010.
- [47] K. S. Camilus, V. K. Govindan, and P. S. Sathidevi, "Computer-aided identification of the pectoral muscle in digitized mammograms," *Journal of Digital Imaging*, vol. 23, pp. 562–580, 2010.
- [48] K. Nie, D. Chang, J.-H. Chen, T.-C. Shih, C.-C. Hsu, O. Nalcioglu, and M.-Y. Su, "Impact of skin removal on quantitative measurement of breast density using MRI," *Med Phys*, vol. 37, pp. 227–233, 2010.
- [49] C. Klifa, J. Carballido-Gamio, L. Wilmes, A. Laprie, J. Shepherd, J. Gibbs, B. Fan, S. Noworolski, and N. Hylton, "Magnetic resonance imaging for secondary assessment of breast density in a high-risk cohort," *Magn Reson Imaging*, vol. 28, pp. 8–15, 2010.
- [50] A. Gubern-Mérida, M. Kallenberg, R. Martí, and N. Karssemeijer, "Fully automatic fibroglandular tissue segmentation in breast MRI: an atlas-based approach," in *MICCAI Workshop: Breast Image Analysis*, 2011.
- [51] L. Wang, K. Filippatos, O. Friman, and H. K. Hahn, "Fully automated segmentation of the pectoralis muscle boundary in breast MR images," in *Medical Imaging*, vol. 7963 of *Proceedings of the SPIE*, 2011.
- [52] C. Gallego and A. L. Martel, "Automatic model-based 3D segmentation of the breast in MRI," in *Medical Imaging*, Proceedings of the SPIE, 2011.
- [53] T. E. Yankeelov, J. J. Luci, M. Lepage, R. Li, L. Debusk, P. C. Lin, R. R. Price, and J. C. Gore, "Quantitative pharmacokinetic analysis of DCE-MRI data without an arterial input function: a reference region model," *Magn Reson Imaging*, vol. 23, pp. 519–529, May 2005.
- [54] P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert, "Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy," *Neuroimage*, vol. 46, pp. 726–738, 2009.
- [55] S. Klein, U. A. van der Heide, I. M. Lips, M. van Vulpen, M. Staring, and J. P. W. Pluim, "Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information," *Med Phys*, vol. 35, pp. 1407–1417, 2008.

- [56] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, “elastix: a toolbox for intensity-based medical image registration,” *IEEE Trans Med Imaging*, vol. 29, pp. 196–205, 2010.
- [57] S. Behrens, H. Laue, M. Althaus, T. Böhler, B. Kuemmerlen, H. K. Hahn, and H.-O. Peitgen, “Computer assistance for MR based diagnosis of breast cancer: Present and future challenges,” *Comput Med Imaging Graph*, vol. 31, pp. 236–247, 2007.
- [58] C. M. Vachon, C. H. van Gils, T. A. Sellers, K. Ghosh, S. Pruthi, K. R. Brandt, and V. S. Pankratz, “Mammographic density, breast cancer risk and risk prediction,” *Breast Cancer Res*, vol. 9, p. 217, 2007.
- [59] M. Khazen, R. M. L. Warren, C. R. M. Boggis, E. C. Bryant, S. Reed, I. Warsi, L. J. Pointon, G. E. Kwan-Lim, D. Thompson, R. Eeles, D. Easton, D. G. Evans, M. O. Leach, and for the Collaborators in the United Kingdom Medical Research Council Magnetic Resonance Imaging in Breast Screening (MARIBS) Study, “A pilot study of compositional analysis of the breast and estimation of breast mammographic density using three-dimensional T1-weighted magnetic resonance imaging,” *Cancer Epidemiol Biomarkers Prev*, vol. 17, pp. 2268–2274, 2008.
- [60] S. van Engeland, S. Timp, and N. Karssemeijer, “Finding corresponding regions of interest in mediolateral oblique and craniocaudal mammographic views,” *Med Phys*, vol. 33, pp. 3203–3212, 2006.
- [61] N. A. Lee, H. Rusinek, J. Weinreb, R. Chandra, H. Toth, C. Singer, and G. Newstead, “Fatty and fibroglandular tissue volumes in the breasts of women 20-83 years old: comparison of x-ray mammography and computer-assisted MR imaging,” *AJR Am J Roentgenol*, vol. 168, pp. 501–506, 1997.
- [62] J. Wei, H.-P. Chan, M. A. Helvie, M. A. Roubidoux, B. Sahiner, L. M. Hadjiiski, C. Zhou, S. Paquerault, T. Chenevert, and M. M. Goodsitt, “Correlation between mammographic density and volumetric fibroglandular tissue estimated on breast MR images,” *Med Phys*, vol. 31, pp. 933–942, 2004.
- [63] S. Wu, S. Weinstein, B. M. Keller, E. F. Conant, and D. Kontos, “Fully-automated fibroglandular tissue segmentation in breast MRI,” in *Proceedings of the 11th international conference on Breast Imaging, IWDM’12*, (Berlin, Heidelberg), pp. 244–251, Springer-Verlag, 2012.
- [64] S. Wu, S. P. Weinstein, E. F. Conant, and D. Kontos, “Automated fibroglandular

-
- tissue segmentation and volumetric density estimation in breast MRI using an atlas-aided fuzzy c-means method.,” *Med Phys*, vol. 40, no. 12, p. 122302, 2013.
- [65] A. Gubern-Mérida, M. Kallenberg, R. Martí, and N. Karssemeijer, “Segmentation of the pectoral muscle in breast MRI using atlas-based approaches,” in *Med Image Comput Comput Assist Interv*, vol. 15 of *Lect Notes Comput Sci*, pp. 371–378, 2012.
- [66] A. Dempster, N. Laird, and D. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *J R Stat Soc*, vol. 39, pp. 1–38, 1977.
- [67] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, “A nonparametric method for automatic correction of intensity nonuniformity in MRI data,” *IEEE Trans Med Imaging*, vol. 17, pp. 87–97, 1998.
- [68] R. Putz and R. Pabst, *Sobotta Atlas of Human Anatomy, Volume 2: trunk, viscera, lower limb*. Philadelphia, PA: Lippincott, Williams and Wilkins, 13th ed., 2001.
- [69] R. Haralick and L. G. Shapiro in *Computer and Robot Vision. Vol. I*, pp. 28–48, Addison-Wesley, 1992.
- [70] N. Otsu, “A threshold selection method from gray level histograms,” *IEEE Trans Syst Man Cybern*, vol. 9, pp. 62–66, 1979.
- [71] F. Ritter, T. Boskamp, A. Homeyer, H. Laue, M. Schwier, F. Link, and H.-O. Peitgen, “Medical image analysis: A visual approach,” *IEEE Pulse*, vol. 2, pp. 60–70, 2011.
- [72] L. R. Dice, “Measures of the amount of ecologic association between species,” *Ecology*, vol. 26, pp. 297–302, 1945.
- [73] A. Klein, J. Andersson, B. A. Ardekani, J. Ashburner, B. Avants, M.-C. Chiang, G. E. Christensen, D. L. Collins, J. Gee, P. Hellier, J. H. Song, M. Jenkinson, C. Lepage, D. Rueckert, P. Thompson, T. Vercauteren, R. P. Woods, J. J. Mann, and R. V. Parsey, “Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration,” *Neuroimage*, vol. 46, pp. 786–802, 2009.
- [74] S. K. Warfield, K. H. Zou, and W. M. Wells, “Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation,” *IEEE Trans Med Imaging*, vol. 23, pp. 903–921, 2004.
- [75] M. Lin, S. Chan, J.-H. Chen, D. Chang, K. Nie, S.-T. Chen, C.-J. Lin, T.-C. Shih, O. Nalcioglu, and M.-Y. Su, “A new bias field correction method combining

- N3 and FCM for improved segmentation of breast density on MRI,” *Med Phys*, vol. 38, pp. 5–14, 2011.
- [76] V. A. McCormack and I. dos Santos Silva, “Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis,” *Cancer Epidemiol Biomarkers Prev*, vol. 15, pp. 1159–1169, 2006.
- [77] C. van Gils, J. D. Otten, A. L. Verbeek, and J. H. Hendriks, “Mammographic breast density and risk of breast cancer: masking bias or causality?,” *Eur J Epidemiol*, vol. 14, pp. 315–320, 1998.
- [78] C. J. D’Orsi, L. W. Bassett, W. A. Berg, S. A. Feig, V. P. Jackson, and D. B. Kopans, *Breast Imaging Reporting and Data System (BI-RADS) Atlas*. 4 ed., 2003.
- [79] J. W. Byng, N. F. Boyd, E. Fishell, R. A. Jong, and M. J. Yaffe, “The quantitative analysis of mammographic densities,” *Phys Med Biol*, vol. 39, pp. 1629–1638, 1994.
- [80] H. Li, M. Giger, O. Olopade, and L. Lan, “Fractal analysis of mammographic parenchymal patterns in breast cancer risk assessment,” *Acad Radiol*, vol. 14, pp. 513–521, 2007.
- [81] M. Nielsen, G. Karemore, M. Loog, J. Raundahl, N. Karssemeijer, J. D. M. Otten, M. A. Karsdal, C. M. Vachon, and C. Christiansen, “A novel and automatic mammographic texture resemblance marker is an independent risk factor for breast cancer,” *Cancer Epidemiol*, vol. 35, pp. 381–387, 2011.
- [82] A. Oliver, X. Lladó, E. Pérez, J. Pont, E. R. E. Denton, J. Freixenet, and J. Martí, “A statistical approach for breast density segmentation,” *J Digit Imaging*, vol. 23, pp. 527–537, 2010.
- [83] A. Torrent, A. Bardera, A. Oliver, J. Freixenet, I. Boada, M. Feixes, R. Marti, X. Llado, J. Pont, E. Perez, S. Pedraza, and J. Marti, “Breast density segmentation: A comparison of clustering and region based techniques,” in *IWDM ’08: Proceedings of the 9th international workshop on Digital Mammography*, (Berlin, Heidelberg), pp. 9–16, Springer-Verlag, 2008.
- [84] K.-H. Ng, C.-H. Yip, and N. A. M. Taib, “Standardisation of clinical breast-density measurement,” *Lancet Oncol*, vol. 13, pp. 334–336, 2012.
- [85] J. A. Shepherd, K. Kerlikowske, L. Ma, F. Duewer, B. Fan, J. Wang, S. Malkov, E. Vittinghoff, and S. R. Cummings, “Volume of mammographic density and risk

-
- of breast cancer,” *Cancer Epidemiol Biomarkers Prev*, vol. 20, pp. 1473–1482, 2011.
- [86] O. Alonzo-Proulx, R. Jong, and M. Yaffe, “Volumetric breast density characteristics as determined from digital mammograms,” *Phys Med Biol*, vol. 57, p. 7443, 2012.
- [87] R. Highnam and M. Brady, *Mammographic Image Analysis*. Kluwer Academic Publishers, 1999.
- [88] J. Kaufhold, J. A. Thomas, J. W. Eberhard, C. E. Galbo, and D. E. G. Trotter, “A calibration approach to glandular tissue composition estimation in digital mammography,” *Med Phys*, vol. 29, pp. 1867–1880, 2002.
- [89] O. Pawluczyk, B. J. Augustine, M. J. Yaffe, D. Rico, J. Yang, G. E. Mawdsley, and N. F. Boyd, “A volumetric method for estimation of breast density on digitized screen-film mammograms,” *Med Phys*, vol. 30, pp. 352–364, 2003.
- [90] S. Ciatto, D. Bernardi, M. Calabrese, M. Durando, M. A. Gentilini, G. Mariscotti, F. Monetti, E. Moriconi, B. Pesce, A. Roselli, C. Stevanin, M. Tapparelli, and N. Houssami, “A first evaluation of breast radiological density assessment by Quantra software as compared to visual classification,” *Breast*, 2012.
- [91] M. Jeffreys, J. Harvey, and R. Highnam, “Comparing a new volumetric breast density method (Volpara™) to Cumulus,” in *IWDM '10: Proceedings of the 10th international workshop on Digital Mammography* (J. Martí, ed.), (Berlin, Heidelberg), pp. 408–413, Springer-Verlag, 2010.
- [92] D. Kontos, P. Bakic, R. J. Acciavatti, E. F. Conant, and A. D. A. Maidment, “A comparative study of volumetric and area-based breast density estimation in digital mammography: results from a screening population,” in *IWDM '10: Proceedings of the 10th international workshop on Digital Mammography*, (Berlin, Heidelberg), pp. 378–385, Springer-Verlag, 2010.
- [93] J. Wang, A. Aziz, D. Newitt, B. N. Joe, N. Hylton, and J. A. Shepherd, “Comparison of Hologic’s Quantra volumetric assessment to MRI breast density,” in *Proceedings of the 11th International Conference on Breast Imaging, IWDM'12*, (Berlin, Heidelberg), pp. 619–626, Springer-Verlag, 2012.
- [94] R. Highnam, N. Sauber, J. H. S. Destounis, and D. McDonald, “Breast density into clinical practice,” in *IWDM '12: Proceedings of the 11th International Workshop on Breast Imaging*, vol. 7361 of *Lect Notes Comput Sci*, pp. 466–473, 2012.

- [95] M. Kallenberg, C. van Gils, M. Lokate, G. den Heeten, and N. Karssemeijer, "Effect of compression paddle tilt correction on volumetric breast density estimation," *Phys Med Biol*, vol. 57, pp. 5155–5168, 2012.
- [96] N. Sauber, A. Chan, and R. Highnam, "BI-RADS breast density classification – an international standard?," in *In European Congress of Radiology: Scientific Exhibit*, 2013.
- [97] F. Li, R. Engelmann, C. E. Metz, K. Doi, and H. MacMahon, "Lung cancers missed on chest radiographs: results obtained with a commercial computer-aided detection program," *Radiology*, vol. 246, pp. 273–80, 2008.
- [98] C. S. White, T. Flukinger, J. Jeudy, and J. J. Chen, "Use of a computer-aided detection system to detect missed lung cancer at chest radiography," *Radiology*, vol. 252, pp. 273–281, 2009.
- [99] S. Kligerman, L. Cai, and C. S. White, "The effect of computer-aided detection on radiologist performance in the detection of lung cancers previously missed on a chest radiograph," *J Thorac Imaging*, vol. 28, pp. 244–252, 2013.
- [100] L. J. W. Burhenne, S. A. Wood, C. J. D’Orsi, S. A. Feig, D. B. Kopans, K. F. O’Shaughnessy, E. A. Sickles, L. Tabar, C. J. Vyborny, and R. A. Castellino, "Potential contribution of computer-aided detection to the sensitivity of screening mammography," *Radiology*, vol. 215, pp. 554–562, 2000.
- [101] J. H. Moltz, L. Bornemann, J.-M. Kuhnigk, V. Dicken, E. Peitgen, S. Meier, H. Bolte, M. Fabel, H.-C. Bauknecht, M. Hittinger, A. Kiessling, M. Pusken, and H.-O. Peitgen, "Advanced segmentation techniques for lung nodules, liver metastases, and enlarged lymph nodes in CT scans," *IEEE J Sel Top Signal Process*, vol. 3, pp. 122–134, 2009.
- [102] B. Platel, R. Mus, T. Welte, N. Karssemeijer, and R. Mann, "Automated characterization of breast lesions imaged with an ultrafast DCE-MR protocol," *IEEE Trans Med Imaging*, pp. 225–232, 2014.
- [103] J. A. Schnabel, D. Rueckert, M. Quist, J. M. Blackall, A. D. Castellano-Smith, T. Hartkens, G. P. Penney, W. A. Hall, H. Liu, C. L. Truwit, F. A. Gerritsen, D. L. G. Hill, and D. J. Hawkes, "A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations," in *Med Image Comput Comput Assist Interv*, vol. LNCS 2208 of *Lect Notes Comput Sci*, pp. 573–581, 2001.

-
- [104] C. Tanner, J. Schnabel, M. Chung, D. Clarkson, D. Rueckert, D. Hill1, and D. Hawkes, “Volume and shape preservation of enhancing lesions when applying non-rigid registration to a time series of contrast enhancing MR breast images,” in *Med Image Comput Comput Assist Interv*, vol. 1935 of *Lect Notes Comput Sci*, pp. 327–337, 2000.
- [105] A. Gubern-Mérida, M. Kallenberg, R. Mann, R. Marti, and N. Karssemeijer, “Breast segmentation and density estimation in breast MRI: a fully automatic framework,” *IEEE J Biomed Health Inform*, vol 19, pp. 349–357, 2015.
- [106] Q. Li, S. Sone, and K. Doi, “Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans,” *Med Phys*, vol. 30, pp. 2040–2051, 2003.
- [107] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, pp. 5–32, 2001.
- [108] K. G. Gilhuijs, M. L. Giger, and U. Bick, “Computerized analysis of breast lesions in three dimensions using dynamic magnetic-resonance imaging,” *Med Phys*, vol. 25, pp. 1647–1654, 1998.
- [109] K. G. A. Gilhuijs, E. E. Deurloo, S. H. Muller, J. L. Peterse, and L. J. S. Kool, “Breast MR imaging in women at increased lifetime risk of breast cancer: clinical system for computerized assessment of breast lesions initial results,” *Radiology*, vol. 225, pp. 907–916, 2002.
- [110] W. Chen, M. L. Giger, U. Bick, and G. M. Newstead, “Automatic identification and classification of characteristic kinetic curves of breast lesions on DCE-MRI,” *Med Phys*, vol. 33, pp. 2878–2887, 2006.
- [111] B. Efron, “Bootstrap methods: Another look at the jackknife,” *Ann Stat*, vol. 7, pp. 1–26, 1979.
- [112] J. Friedman, T. Hastie, and R. Tibshirani, “Special invited paper. Additive logistic regression: A statistical view of boosting,” *Ann Stat*, vol. 28, pp. 337–374, 2000.
- [113] C.-C. Chang and C.-J. Lin, *LIBSVM: a library for support vector machines*, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [114] T.-F. Wu, C.-J. Lin, and R. C. Weng, “Probability estimates for multi-class classification by pairwise coupling,” *J Mach Learn Res*, vol. 5, pp. 975–1005, Dec. 2004.

- [115] E. W. L. Chuwa, A. W. Y. Yeo, H. N. Koong, C. Y. Wong, W. S. Yong, P. H. Tan, J. T. S. Ho, J. S. L. Wong, and G. H. Ho, “Early detection of breast cancer through population-based mammographic screening in asian women: a comparison study between screen-detected and symptomatic breast cancers,” *Breast J*, vol. 15, no. 2, pp. 133–139, 2009.
- [116] R. M. Mann, J. Veltman, H. Huisman, and C. Boetes, “Comparison of enhancement characteristics between invasive lobular carcinoma and invasive ductal carcinoma,” *J Magn Reson Imaging*, vol. 34, pp. 293–300, 2011.
- [117] G. D. Leonard and S. M. Swain, “Ductal carcinoma in situ, complexities and challenges,” *J Natl Cancer Inst*, vol. 96, pp. 906–920, Jun 2004.
- [118] C. K. Kuhl, “The ”coming of age” of nonmammographic screening for breast cancer,” *JAMA*, vol. 299, pp. 2203–2205, 2008.
- [119] J. Levman, T. Leung, P. Causer, D. Plewes, and A. L. Martel, “Classification of dynamic contrast-enhanced magnetic resonance breast lesions by support vector machines,” *IEEE Trans Med Imaging*, vol. 27, pp. 688–696, 2008.
- [120] Y.-C. Chang, Y.-H. Huang, C.-S. Huang, P.-K. Chang, J.-H. Chen, and R.-F. Chang, “Classification of breast mass lesions using model-based analysis of the characteristic kinetic curve derived from fuzzy c-means clustering,” *J Magn Reson Imaging*, vol. 30, pp. 312–322, Apr 2012.
- [121] W. Chen, M. L. Giger, and U. Bick, “A fuzzy c-means (FCM)-based approach for computerized segmentation of breast lesions in dynamic contrast-enhanced MR images,” *Acad Radiol*, vol. 13, pp. 63–72, 2006.
- [122] A. B. Ashraf, S. C. Gavenonis, D. Daye, C. Mies, M. A. Rosen, and D. Kontos, “A multichannel markov random field framework for tumor segmentation with an application to classification of gene expression-based breast cancer recurrence risk,” *IEEE Trans Med Imaging*, vol. 32, pp. 637–648, 2013.
- [123] T. Tan, B. Platel, H. Huisman, C. I. Sánchez, R. Mus, and N. Karssemeijer, “Computer aided lesion diagnosis in automated 3D breast ultrasound using coronal spiculation,” *IEEE Trans Med Imaging*, vol. 31, pp. 1034–1042, 2012.
- [124] A. Gubern-Mérida, L. Wang, M. Kallenberg, R. Martí, H. Hahn, and N. Karssemeijer, “Breast segmentation in MRI: quantitative evaluation of three methods,” in *Medical Imaging*, Proceedings of the SPIE, pp. 86693G–86693G–7, 2013.

-
- [125] M. O. Leach, A. K. Boggis, C. R. Dixon, D. F. Easton, R. A. Eeles, D. G. Evans, F. J. Gilbert, I. Griebisch, R. J. Hoff, P. Kessar, S. R. Lakhani, S. M. Moss, A. Nerurkar, A. R. Padhani, L. J. Pointon, D. Thompson, R. M. Warren, and MARIBS study group., “Screening with magnetic resonance imaging and mammography of a UK population at high familial risk of breast cancer: a prospective multicentre cohort study (MARIBS),” *Lancet*, vol. 365, pp. 1769–78, 2005.
- [126] C. Kuhl, S. Schrading, S. Weigel, K. Nüssle-Kügele, H. Sittek, B. Arand, N. Morakkabati, C. Leutner, B. Tombach, D. Nordhoff, C. Perlet, A. Rieber, W. Heindel, H. Brambs, and H. Schild, “[the ”EVA” trial: Evaluation of the efficacy of diagnostic methods (mammography, ultrasound, MRI) in the secondary and tertiary prevention of familial breast cancer. Preliminary results after the first half of the study period],” *RoFo : Fortschritte auf dem Gebiete der Rontgenstrahlen und der Nuklearmedizin*, vol. 177, p. 818—827, June 2005.
- [127] F. Sardanelli, F. Podo, G. D’Agnolo, A. Verdecchia, M. Santaquilani, R. Musumeci, G. Trecate, S. Manoukian, S. Morassut, C. de Giacomi, M. Federico, L. Cortesi, S. Corcione, S. Cirillo, and V. Marra, “Multicenter comparative multimodality surveillance of women at genetic-familial high risk for breast cancer (HIBCRIT study): Interim results,” *Radiology*, vol. 242, no. 3, pp. 698–715, 2007.
- [128] F. Sardanelli, F. Podo, F. Santoro, S. Manoukian, S. Bergonzi, G. Trecate, D. Vergnaghi, M. Federico, L. Cortesi, S. Corcione, S. Morassut, C. Di Maggio, A. Cilotti, L. Martincich, M. Calabrese, C. Zuiani, L. Preda, B. Bonanni, L. A. Carbonaro, A. Contegiacomo, P. Panizza, E. Di Cesare, A. Savarese, M. Crecco, D. Turchetti, M. Tonutti, P. Belli, A. D. Maschio; High Breast Cancer Risk Italian 1 (HIBCRIT-1) Study, “Multicenter surveillance of women at high genetic breast cancer risk using mammography, ultrasonography, and contrast-enhanced magnetic resonance imaging (the high breast cancer risk italian 1 study),” *Invest Radiol*, vol. 46, no. 2, pp. 94–105, 2011.
- [129] J. A. van Dijck, A. L. Verbeek, J. H. Hendriks, and R. Holland, “The current detectability of breast cancer in a mammographic screening program. A review of the previous mammograms of interval and screen-detected cancers,” *Cancer*, vol. 72, pp. 1933–1938, 1993.
- [130] J. A. Harvey, L. L. Fajardo, and C. A. Innis, “Previous mammograms in patients with impalpable breast carcinoma: retrospective vs blinded interpretation. 1993 ARRS President’s Award,” *AJR Am J Roentgenol*, vol. 161, pp. 1167–1172, 1993.

- [131] R. D. Jones, L. McLean, J. R. Young, W. Simpson, and F. Neilson, "Proportion of cancers detected at the first incident screen which were false negative at the prevalent screen," *Breast*, vol. 5, pp. 339–343, 1996.
- [132] R. G. Blanks, M. G. Wallis, and S. M. Moss, "A comparison of cancer detection rates achieved by breast cancer screening programmes by number of readers, for one and two view mammography: results from the UK national health service breast screening programme," *J Med Screen*, vol. 5, pp. 195–201, 1998.
- [133] B. Vitak, "Invasive interval cancers in the Östergötland mammographic screening programme: radiological analysis," *Eur Radiol*, vol. 8, pp. 639–646, 1998.
- [134] N. Karssemeijer, J. D. Otten, A. L. Verbeek, J. H. Groenewoud, H. J. de Koning, J. H. Hendriks, and R. Holland, "Computer-aided detection versus independent double reading of masses on mammograms," *Radiology*, vol. 227, pp. 192–200, 2003.
- [135] A. Malich, D. R. Fischer, and J. Böttcher, "CAD for mammography: the technique, results, current role and further developments," *Eur Radiol*, vol. 16, pp. 1449–1460, 2006.
- [136] S. Leon, L. Brateman, J. Honeyman-Buck, and J. Marshall, "Comparison of two commercial CAD systems for digital mammography," *J Digit Imaging*, vol. 22, pp. 421–423, 2009.
- [137] S. Schalekamp, B. van Ginneken, N. Karssemeijer, and C. M. Schaefer-Prokop, "Chest radiography: new technological developments and their applications," *Seminars in Respiratory & Critical Care Medicine*, vol. 35, pp. 3–16, 2014.
- [138] L. Gallardo-Estrella, B. van Ginneken, and E. M. van Rikxoort, "Normalization of CT scans reconstructed with different kernels to reduce variability in emphysema measurements," in *Medical Imaging*, vol. 8670 of *Proceedings of the SPIE*, p. 86700E, 2013.
- [139] R. H. El Khouli, K. J. Macura, I. R. Kamel, M. A. Jacobs, and D. A. Bluemke, "3-T Dynamic contrast-enhanced MRI of the breast: Pharmacokinetic parameters versus conventional kinetic curve analysis," *American Journal of Roentgenology*, vol. 197, pp. 1498–1505, Dec. 2011.
- [140] R. M. Mann, R. D. Mus, J. van Zelst, C. Geppert, N. Karssemeijer, and B. Platel, "A novel approach to contrast-enhanced breast magnetic resonance imaging for screening: High-resolution ultrafast dynamic imaging," *Invest Radiol*, 2014.

- [141] T. Mertzaniidou, J. Hipwell, M. J. Cardoso, X. Zhang, C. Tanner, S. Ourselin, U. Bick, H. Huisman, N. Karssemeijer, and D. Hawkes, "MRI to x-ray mammography registration using a volume-preserving affine transformation," *Med Image Anal*, vol. 16, pp. 966–975, 2012.
- [142] G. Abramovici and M. B. Mainiero, "Screening breast MR imaging: Comparison of interpretation of baseline and annual follow-up studies," *Radiology*, vol. 259, no. 1, pp. 85–91, 2011.
- [143] T. C. Williams, W. B. DeMartini, S. C. Partridge, S. Peacock, and C. D. Lehman, "Breast MR imaging: computer-aided evaluation program for discriminating benign from malignant lesions," *Radiology*, vol. 244, pp. 94–103, 2007.
- [144] D. B. Kopans, "Basic physics and doubts about relationship between mammographically determined tissue density and breast cancer risk," *Radiology*, vol. 246, no. 2, pp. 348–353, 2008.

Acknowledgments

The story of my PhD started at the end of 2009. I was studying for my master's degree in Girona and working on a medical image analysis project at the VICOROB group of the University of Girona. One day I walked into Robert's office (Robert is my co-promotor at the University of Girona) and told him that I wanted to do my master's thesis abroad. A few months later (beginning 2010), I started a 4 months stay, under the supervision of prof. dr. Nico Karssemeijer (my promotor), in Nijmegen. At that time, I only knew that the topic of my research was "making a computer do something with breast MR images". Working on that topic took longer than I expected, but, five years later, and after many bi-directional trips between Girona and Nijmegen, it seems that I might be able to get a PhD degree out of it.

Prof. dr. Karssemeijer, beste Nico, I would like to thank you for opening the doors of DIAG to me, accepting me as a visiting master student back in 2009 and later on giving me the opportunity to pursue a PhD in the Netherlands. You provided me excellent supervision. Furthermore, thanks to your knowledge on the topic, your ability to seeing things from the adequate perspective and your preference for exactness, my personal and professional skills improved substantially. Dr. Martí, benvolgut Robert, estic molt content d'haver-te tingut com a supervisor. Moltes gràcies per introduir-me en aquest món de la recerca quan tot just estava en el 3r any de carrera. He après molt de tu i m'ho he passat molt bé treballant plegats. Sobretot, t'agraeixo la flexibilitat i disponibilitat que has tingut durant tots aquests anys. Dr. Platel, beste Bram, many thanks for being my co-promotor and helping me with the last research ideas and writing papers. I had a lot of fun discussing and working on it with you. Your contribution has increased the quality of my thesis significantly. Dr. Mann, beste Ritse, thank you for your clinical insight and perspective. It has been really useful to be better able to understand what my research could do for clinical practice.

Prof. dr. Tom Heskes, prof. dr. Josien Pluim and prof. sir. Michael Brady, I would like to thank you for participating in the reading committee.

I want to thank all the people I met at the Diagnostic Image Analysis Group (DIAG) in Nijmegen. I have really enjoyed the time spent there. It is a great place to do research. One can learn a lot from all the regular meetings and discussions going on about the different research topics. Furthermore, I have had the pleasure to meet many friendly people. This good atmosphere has been also translated outside the office in all kind of social activities such as football matches, watching football, drinks (while talking about football) and, not less important, the memorable DIAG weekends. Michiel, I am really grateful for the warm welcome you gave me as soon as I stepped into the barracks. You helped me a lot getting to know the group and introducing me to the Dutch culture (Cruesli included). I really enjoyed working with you and having you around. Tao, it has been a pleasure sharing time together at DIAG. I had a lot of fun

with you in Nijmegen and at conferences. Furthermore, being involved in your beloved ABUS topic has made me realize even more what a good researcher you are. Jan-Jurre, I am fortunate to have you as my paranimf. You bring a really good vibe into the group and I am happy that I do not only benefit from that at work (although I also think that we should go out for a drink more often). Thanks also to these DIAG ladies with whom, for the time being, I share nationality (but not the same feeling for it). Clarisa y Leti, es un placer poder desconectar del mundo holandés de vez en cuando con vosotras. Also with the same nationality (and I think that we also share the feeling for it), Jaime, gracias por tu ayuda en la parte final del doctorado, en especial con todo lo relacionado con PR. Y no menos importante, gracias también por ayudarme con la botella de tequila que tengo en casa. Suzan, it is great that you have joined DIAG. We all benefit from your never-ending cheerfulness. Thanks for helping me with the last chapter of the thesis. Finally, the names of Rick, Mark, Wendy, Pragnya, Geert, Colin, Steven, Thijs and Jan also deserve to appear in these pages. All of you have been, on your own way, a good support during this time and I really enjoyed working with you around.

Many thanks to people at the VICOROB group from the Universitat de Girona. Gràcies per tots els cafès, dinars, sopars, reunions, conferències i simposiums. Joseta i Mireia, gràcies per la vostra ajuda i paciència per arreglar tota la paperassa generada per les meves estades, beques i viatges. Gràcies també als que treballen en mammo, Yago, Arnau i Joan, amb els quals ha estat un plaer col·laborar-hi.

I also want to thank my colleagues at Fraunhofer Mevis Lei, Fabian and Hendrik, and the ASSURE mates Julia and Oliver. Our regular ASSURE TCON meetings have been really useful to think and discuss about possible applications and improvements of the methods developed in this PhD.

A word of thanks should go to the guys I met while studying at the University of Girona. During our respective BSc, MSc and PhD studies, we saw each other almost daily until I moved to the Netherlands. Thank you Gerard, Pablo, Ricard, Enric, Massi, Eloy, Mariano and Albert (aka Plà). It has been a great pleasure sharing all this time together while we were studying, working, traveling, talking, eating, drinking, watching Lost, playing StrikerManager and, last but not least, celebrating F.C. Barcelona's most successful period of its history. A tots vosaltres us agraeixo fer temps per mi i fer-me sentir com si mai hagués marxat a l'estranger cada vegada que vinc per Girona. Gerard, un plaer haver estat companys de lab durant molt de temps. No ens ha servit per millorar a nivell professional, però sí per mantenir-nos al dia sobre les nostres intimitats, i molt més important, discutir sobre el Barça. De fet, ara que hi penso, el nivell de comunicació entre nosaltres no ha canviat gaire gràcies a google. Malauradament, el nivell del Barça sí. Albert, és un honor per mi tenir-te com a paranimf. Un lamentable

PCFútbol ens va unir i les apostes en partits de la lliga txeca (i per *poles* de'n Trulli!) van acabar de confirmar que tenim moltes coses en comú. Crec que fèiem un molt bon equip durant totes les pràctiques que hem dut a terme plegats en la carrera. A veure si en un futur trobem un projecte en comú on puguem explotar-ho.

Thanks to the friends I met in Nijmegen Ruben, Luc and Tim, and to my other friends from Girona that, although not seeing them as often as I would like, I had their support and shared really good moments during this time. Albert (Delgado), gràcies per les teves visites. És genial sentir, cada vegada que ens trobem, com si encara fòssim aquells *amigos* de 3 anys. Raul, tan de bo tothom aprengué del teu positivisme. Explicar-nos les nostres aventures a l'estranger i visitar-nos força sovint ho ha fet tot més fàcil i divertit.

Special thanks to my families. Saskia, Bas, Joost, Oma, Peter and Nettie, heel erg bedankt voor het verwelkomen van mij in de *don munstergelene* familie. Mama, tu tirar adelante a pesar de los malos momentos vividos me ha servido de ejemplo para también tirar adelante esta tesi. Tata, Jordi i Núria, tot i viure lluny de vosaltres sento el vostre suport i, per això, us estic molt agraït.

Lieve Dymphie. I truly believe that without your support, understanding and patience, this thesis would have been more difficult to accomplish. Thank you. I am really happy to be with you (and Nino).

Finally, gràcies Papa per tot l'esforç i energia que vas dedicar per tal de quedar-te entre nosaltres. M'entristeix molt moltíssim no haver pogut compartir tot això amb tu. Crec que n'haguessis gaudit tan o més que jo. No obstant, vull que sàpigues que la teva lluita ha estat una gran motivació i que aquest èxit és, sense cap mena de dubte, també tot teu.

Curriculum Vitae

Curriculum Vitae



Albert Gubern-Mérida was born in Girona, Catalonia, Spain, on February 19th 1985. He studied Computer Science Engineering at the University of Girona (UdG). He carried out his MSc thesis on “Multi-class probabilistic atlas-based segmentation in breast MR images” in collaboration with the Diagnostic Image Analysis Group (DIAG) in Nijmegen. As of October 2010, he worked as a PhD student at the VICOROB group (UdG) and at DIAG. His work on automated analysis of breast MRI is described in this thesis.