Automatic Segmentation and Classification Of Breast Lesions Using a Novel Multimodal Imaging Approach

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Visual Computing

eingereicht von

Wolf-Dieter Vogl, BSc

Matrikelnummer 0626355

an der
Fakultät für Informatik der Technischen Universität Wien

Betreuung: a.o.Univ.-Prof. Dipl.-Ing. Dr.techn. Robert Sablatnig
Mitwirkung: Ass.Prof. Dipl.-Ing. Dr. Georg Langs

Wien, 6. Mai 2012

_____   _____
(Unterschrift Verfasser)   (Unterschrift Betreuung)

Technische Universität Wien
A-1040 Wien ▪ Karlsplatz 13 ▪ Tel. +43-1-58801-0 ▪ www.tuwien.ac.at

Erklärung zur Verfassung der Arbeit

Wolf-Dieter Vogl, BSc
Vorgartenstraße 207/2/21, 1020 Wien

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

_____   _____

(Ort, Datum)   (Unterschrift Verfasser)

i

Acknowledgements

I want to express my gratitude to all persons supporting me in writing this thesis. First and foremost I want to thank Georg Langs from the CIR lab at the Medical University of Vienna, who guided me through the whole work, from the first experiments to the proof reading of the thesis. Without his constant input and constructive comments on my work it would not be of this quality and size.

I want to thank Thomas Helbich from the Department of Radiology at the Medical University of Vienna for supporting my work and funding the thesis. He is also head of the multimodal breast lesion study, on which this thesis is founded. Furthermore, I want to thank Katja Pinker and Hubert Bickel from the Department of Radiology for their constant support in all medical issues, such as multimodal data acquisition, lesion segmentation, and reflection of the multimodal approach from a medical viewpoint.

My gratitude also goes to Günther Grabner from the MR Centre of Excellence at the Medical University of Vienna for granting an facilitating an internship at the Centre, which aroused my interest in medical image processing. He also enabled the opportunity of making this Master's thesis.

I will also thank my supervisor Robert Sablatnig from the Computer Vision Lab (CVL) at the Vienna University of Technology for reviewing and commenting the thesis. The lessons learned in the "Scientific Presentation and Communication" seminar lectured by him and Amy Krois-Lindner have been directly applied on writing and presenting the thesis, causing an improvement in science and language quality.

Special thanks goes to my parents and my family, who supported me all over the years in my studies. Finally, I want to say 'Thank You' to Tanja, my wonderful partner, who was constantly backing me all the way through the studies and the Master's thesis.

Abstract

Medical imaging has become crucial in detection and diagnosis of breast cancer. Advanced image modalities like Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI), Diffusion Weighted Magnetic Resonance Imaging (DWI), Positron Emission Tomography (PET), and Magnetic Resonance Spectroscopy (MRS) provide complementary information about the lesion biology and increase the precision and certainty of cancer diagnosis. A systematic analysis and categorization of the image data is essential for an accurate diagnosis. Computer Aided Diagnosis (CAD) systems aid radiologists in this task by using digital image analysis methods and machine learning algorithms.

The aim of this thesis is the development of a novel multimodal breast lesion CAD system with a fully automatic combination of several medical image modalities, the fully automatic detection and segmentation of cancerous regions, and the fully automatic classification of the lesions into benign and malign ones. As modalities DCE-MRI, DWI and PET are used in this initial project, with the opportunity to add additional modalities later on. The CAD system consists of three main elements: the registration of the image modalities using a Large Deformation Diffeomorphic Metric Mapping (LDDMM) method, the detection and segmentation of the lesion using a Random Forests (RF) algorithm, and the classification of the lesions into benign and malign using a further RF.

In the validation process it is qualitatively and quantitatively demonstrated on 16 breast studies that a multimodal approach improves the segmentation and classification performance in comparison to a single-modal DCE-MRI approach. The segmentation performance measured by the mean Dice Similarity Coefficient (DSC) increased from 0.39 to 0.45. The classification performance measured by Sensitivity / Specificity increased from 61% / 68% to 89% / 99%, and the Area under Curve (AuC) of a Receiver Operating Characteristic (ROC) curve increased from 0.65 to 0.94. The results of the validation also demonstrate that the proposed method provides a comparable performance to state-of-the-art single-modality CAD systems.

Kurzfassung

Ein wesentlicher Bestandteil für die Erkennung und Diagnose von Brustkrebs sind medizinische Bildgebungsverfahren. Moderne Bildgebungsverfahren wie Dynamische Kontrasterweiterte Magnetresonanztomographie (DCE-MRI), Positronemissionstomographie (PET), Diffusionsgewichtetes MRT (DWI) und Magnetresonanzspektroskopie (MRS) liefern eine Vielzahl an Informationen über die Tumorbiologie und ermöglichen somit eine genauere und sicherere Diagnose von Brustkrebs. Eine systematische Analyse und Kategorisierung der Bilddaten ist ein essentieller Bestandteil einer präzisen Diagnose. Computerunterstützte Diagnosesysteme (CAD) unterstützen Radiologen bei der Analyse der Bilddaten durch den Einsatz von digitalen Bildanalysemethoden und Maschinenlernalgorithmen.

In dieser Arbeit wird ein neuartiges multimodales CAD System beschrieben, welches eine automatische Kombination der Bildmodalitäten, eine automatische Lokalisierung und Segmentierung von Brustkrebsgewebe, und eine automatische Klassifizierung der Läsionen in benign und malign ermöglicht. Die verwendeten Modalitäten in diesem Projekt sind DCE-MRI, PET und DWI, wobei weitere Modalitäten später hinzugefügt werden können. Das CAD System besteht aus drei Hauptteilen: die Registrierung der Bildmodalitäten mit Hilfe einer Large Deformation Diffeomorphic Metric Mapping (LDDMM) Methode, der Detektion und Segmentierung der Läsion durch einen Random Forests (RF) Klassifikator, und der Unterscheidung von benignen und malignen Läsionen durch einen weiteren RF Klassifikator.

Im Validierungsprozess konnte anhand von 16 Patientendaten gezeigt werden, dass der multimodale Ansatz sowohl die Segmentierung als auch die Klassifizierung verbessert, verglichen mit einem Ansatz basierend auf der DCE-MRI Modalität. Die Segmentierungsgenauigkeit, gemessen durch den mittleren Dice Similarity Coefficient (DSC), stieg von 0.39 auf 0.45. Die Sensitivität / Spezifität der Klassifizierung stieg von 61 % / 68 % auf 89 % / 99 % und Area under Curve (AuC) der Receiver Operating Characteristic (ROC) stieg von 0.65 auf 0.94. Die Ergebnisse der Validierung zeigen auch, dass dieser Ansatz vergleichbare Resultate zu aktuellen unimodalen CAD Systemen bietet.

Contents

Acronyms

AC	Active Contour
ADC	Apparent Diffusion Coefficient
AuC	Area under Curve
CA	Computational Anatomy
CAD	Computer Aided Detection & Diagnosis
CADe	Computer Aided Detection
CADx	Computer Aided Diagnosis
CART	Classification And Regression Tree
CC	Cross-Correlation
CKC	Characteristic Kinetic Curve
CT	Computed Tomography
DCE-MRI	Dynamic Contrast Enhanced Magnetic Resonance Imaging
DD	Diffeomorphic Demons
DSC	Dice Similarity Coefficient
DTI	Diffusion Tensor Imaging
DWI	Diffusion Weighted magnetic resonance Imaging
FCM	Fuzzy C-Means
FDG	FluoroDesoxyGlucose
FFD	Free-Form Deformation
GLCM	Gray-Level Co-occurrence Matrix
GrSyN	Greedy Symmetric Normalization
LDDMM	Large Deformation Diffeomorphic Metric Mapping
LOOCV	Leave-One-Out Cross-Validation
MaxEnh	Maximum of intensity Enhancement
MaxSlope	Maximum of Slope
MDER	Maximum of DERivative
MI	Mutual Information
MRI	Magnetic Resonance Imaging
MRF	Markov Random Field
mRMR	minimum-Redundancy-Maximum-Relevance
MRS	Magnetic Resonance Spectroscopy
MSQ	Mean Squared intensity difference
NIfTI	Neuroimaging Informatics Technology Initiative
OOB	Out-Of-Bag

PCA   Principal Component Analysis
PET   Positron Emission Tomography
RF   Random Forests
ROC   Receiver Operating Characteristic
ROI   Region Of Interest
SNR   Signal-to-Noise Ratio
SSD   Sum of Squared Differences
SVM   Support Vector Machine
SyN   Symmetric Normalization
SyGN   Symmetric Geodesic Normalization
THM   Time to Half Maximum
TPS   Thin-Plate Spline
TTP   Time To Peak

Introduction

Breast cancer is the most common cancer among women, and a leading cause of cancer death in most Western countries [53]. Early detection and a precise diagnosis are essential for efficient treatment [18]. Radiological imaging, like mammography [107], or breast ultrasound [52] are used for detection and biopsy of suspicious lesions. These two modalities are clinical standards for breast cancer detection and characterization. However, both modalities are still limited in lesion detection with regard to both sensitivity and specificity [83, 163]. In addition, mammography suffers from a high false positive rate of 16.3% at first and 9.6% at subsequent mammography [75]. Alternative imaging methods such as Magnetic Resonance Imaging (MRI) [164] or Positron Emission Tomography (PET) [182] have been developed with the aim of overcoming the limitations of mammography and ultrasound. MRI is reported on the one hand to be a sensitive method; on the other hand a low to moderate specificity is reported [85]. PET provides functional information through metabolism about the lesion; hence it suffers from a low spatial resolution and a lack of morphological information, resulting in a difficult localization of the lesion [153]. To overcome this limitation hybrid PET / Computed Tomography (CT) scanner [147] and recently PET / MRI scanners [5, 78] have been developed, providing a combination of functional and morphological information about the lesion. MRI-based functional and metabolic imaging methods such as Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI) [119], Diffusion Weighted magnetic resonance Imaging (DWI) [24], and Magnetic Resonance Spectroscopy (MRS) [63] have been shown to provide an additional diagnostic value to the morphological MRI modality [63, 87, 97].

Combining these multimodal imaging methods is a novel and promising research area, where an increase in the accuracy of cancer diagnosis, assessment and therapy progress is expected [56, 117, 118].

To aid radiologists in detection and diagnosis of breast cancer, Computer Aided Detection & Diagnosis (CAD) systems have been proposed [15]. Such tools are able to analyze a large amount of image data in reasonable time, detect and visualize complex correlations and patterns, as well as provide objective and repeatable results [168], yielding in an increased accuracy of a diagnosis in many cases [88]. CAD systems are divided into two categories: Computer Aided

Detection (CADe) and Computer Aided Diagnosis (CADx). CADe systems assist radiologists in localizing suspicious regions in medical images, whereas CADx systems support the radiologist in the diagnosis of suspicious regions by providing and analyzing information extracted from these regions [159].

The aim of this thesis is the development of a breast lesion CAD system with a fully automatic combination of several medical image modalities, the fully automatic detection and segmentation of cancerous regions, and the fully automatic classification of the lesions into benign and malign ones. As modalities DCE-MRI, DWI and PET are used in this initial project with the opportunity to add additional modalities later on. The performance of the CAD system is evaluated and compared with single modality approaches.

1.1 Problem Statement

In this thesis it is hypothesized that the various breast image modalities provide complementary information, resulting in an improved performance when combining them in a breast lesion CAD system. Thus, the aim of this thesis is the development of a CAD framework that combines and interprets the 3-dimensional image modalities PET, MRI, DCE-MRI, and DWI. These modalities give insight into morphological changes (MRI) as well as changes at the cellular level such as angiogenesis (DCE-MRI), metabolism (PET) and cell function (DWI) of cancerous and non-cancerous tissue. In this thesis it is evaluated quantitatively, which combinations of image modalities are suitable for an accurate localization of the lesion and for an accurate cancer diagnosis, and what the benefit of each modality is in a CAD system.

A crucial step in such a framework is the accurate fusion of the image modalities. In this fusion process deformations of the breast, introduced by varying patient positions in the scanners and image distortions of the DWI modality [74], have to be compensated. The fused data are then used for automatic detection and segmentation of cancerous tissue (CADe), as well as the automatic classification of the lesions into benign and malignant (CADx). The performance of the CAD system has to be validated and compared with state-of-the-art single modality CAD systems.

1.2 Methodological Approach

The CAD system is implemented as a pipeline with three main steps (Figure 1.1):

1. Fusion of the image modalities by *registration* of the 3D images,

2. *Segmentation* of the lesion, and

3. *Classification* of the lesion with regard to being benign or malign.

In the registration step all modalities are aligned by finding correspondences in the images and applying a deforming transformation on them. The segmentation step is implemented as a voxel-based binary classification process, where a Random Forests (RF) machine learning
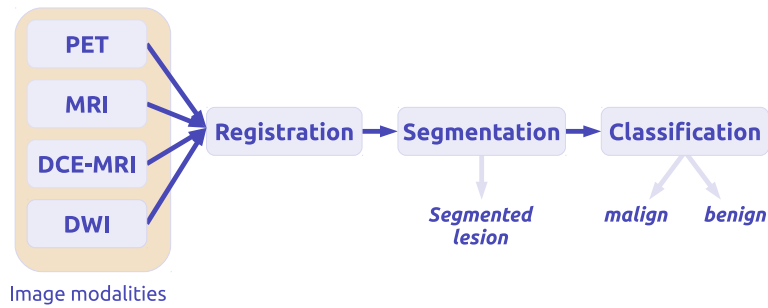
Figure 1.1: The CAD pipeline.

algorithm [28] predicts for each voxel (=3-dimensional pixel) if it is classified as cancerous tissue or normal tissue. Discriminative segmentation features are extracted from the multimodal dataset for each voxel. Manual segmentations annotated by an experienced radiologist are used as ground-truth for validation and classifier training. In addition to classification, the RF algorithm is also used for feature selection, by determining the feature combinations best suited for discriminating between cancerous and normal tissues. The predictive power of the selected features is evaluated using cross-validation. In the classification step a second RF classifier is used to discriminate between malignant and benign lesions. For this reason a multimodal set of features is extracted from the previously obtained automatic segmentation. Instead of using voxel-based features, as for the segmentation, the data from the voxels are subsumed into *clusters* that characterize the entire lesion. These cluster-based features are used for classifying the segmented lesions into malign or benign after identifying relevant features. The classification performance is evaluated in a cross-validation process. The histopathological analysis serves as ground-truth for the training and validation of the RF classifier.

1.3   Contribution of the Thesis

The contribution of this thesis is the development and evaluation of a fully-automatic CADe and CADx framework using a multimodal medical image dataset to localize, segment and classify breast lesions.

Numerous CAD systems based on DCE-MRI are available (see Chapter 4 - 'Related Work'). Although, the usage of a combination of several modalities in a CAD system is a novel approach. The following are the main contributions:

- Automatic fusion of multimodal breast image data (PET, MRI, and DWI) by a registration process,

- Using the fused information from all modalities for automatic segmentation and classification of the lesion,

- Using an advanced machine learning algorithm, which is able to find significant information for segmentation and classification in the multimodal dataset, and

- Using a clustering method for automatic selection of significant regions in DWI and PET modalities for lesion classification.

1.4   Thesis Outline

The thesis is divided into seven chapters. The outline of the thesis is as follows:

1. Introduction. The first chapter provides a coarse overview of the motivation and problem statement of the proposed multimodal breast lesion CAD system. The chapter is closed by a short outline of the medical image modalities used in the CAD pipeline.

2. Multimodal Image Registration. The theoretical background and the general state-of-the-art in medical image registration is provided in this chapter. The focus of this chapter is on non-rigid registration methods, which are able to handle large deformations as well as images of various modalities.

3. Classification and Regression by Random Forests. The functional principle of the RF algorithm is outlined in this chapter. In addition, methods for feature analysis, feature selection and performance evaluation using RF are discussed.

4. Related Work. Related work done by other research groups is outlined in this chapter, with the aim of giving a state-of-the-art overview in multimodal breast image registration, segmentation of breast lesions, and the classification of breast lesions. In addition, a summary of the performance of breast CADx methods is given in this chapter.

5. Methodology. In this chapter the CAD pipeline is presented in detail, from data acquisition, registration, and segmentation to classification. The chapter also contains a detailed description of the features extracted from the images, which are needed for the segmentation and classification.

6. Experiments and Results. This chapter is divided into three main sections, covering the validation and evaluation of the registration process, the segmentation, and the classification. In the registration section several transformation models and similarity metrics are evaluated, in order to optimize the multimodal breast image alignment. In the second section the performance of the segmentation features are examined, with the focus on exploration of the benefit of each modality. The performance of the segmentation using various combinations of features is evaluated using cross-validation. The last section of this chapter presents the examination of the classification features, and their performance in a classification process. The performance and the predictive power of the RF are also evaluated in a cross-validation process.

7. Summary and Conclusion. Finally, in this chapter the essential parts of the thesis are recapitulated, the main results of the experiments are summarized, discussed and compared with the related work, and closing remarks are given. The chapter closes with thoughts about future work and improvements of the CAD system.
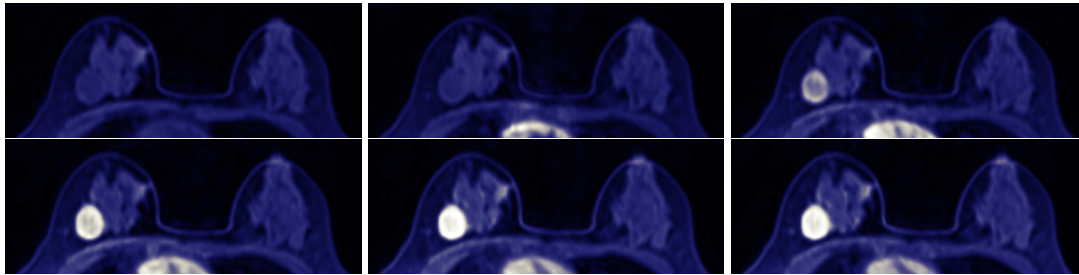
Figure 1.2: Series of DCE-MRI images. The contrast agent accumulates in the cancerous region over time and causes a signal intensity change in the images. The chronological order of the images is from left to right and top to bottom.

1.5 Medical Image Modalities

In this section a short overview of the modalities used in this thesis is given. It describes the information each modality provides, the contribution of the modality in a diagnosis system, and the performance of the modality in breast cancer diagnosis.

1.5.1 Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI)

DCE-MRI is used amongst other modalities for localizing breast lesions and differentiation of benign from malignant breast lesions by analyzing the special microvascular structure of cancerous tumors [85]. *Angiogenesis* is the process of growing new blood vessels from existing vessels [134]. At normal tissue this process stops when the blood supply of the tissue is sufficient. Invasive cancers keep the angiogenesis active to satisfy their increasing metabolic demand for oxygen and nutrients, resulting in a growth of exponential formation of microvascular structures in the cancerous tissue.

To get a DCE-MRI series a contrast agent is intravenously applied, while MRI scans are acquired at regular intervals. The contrast agent increases the local magnetization, resulting in higher signal intensities. The dense microvascular structure of cancerous tissue does also have an increased vascular leakage. Therefore, more contrast agent moves from the vascular system to the tumor tissue and is accumulated there. The accumulation causes a rise of the signal intensity in cancerous tissue (Figure 1.2). This change of signal intensity is denoted in the literature as *contrast enhancement* [87] or *enhancement kinetics* [85]. The characteristics of enhancement kinetics in cancerous regions can be used to differ benign from malignant breast lesions [87]. In order to do this a radiologist draws a Region Of Interest (ROI) in the signal enhancing part of the lesion. From this ROI a DCE-MRI *time-signal curve* (also referred to as *kinetic curve*) is computed by averaging the signal intensities in the ROI for each image acquisition timepoint and plotting these values in a diagram (Figure 1.3 left). In addition to a morphologic categorization of the lesion, the characteristic shape of the time-signal intensity curve is used for differential diagnostics, by analyzing the steepness of the contrast uptake in the early phase after applying the contrast agent, and the behavior of the curve in the later phase: whether the signal intensity continues to increase after the initial uptake (Type I), whether it exhibits a plateau (Type
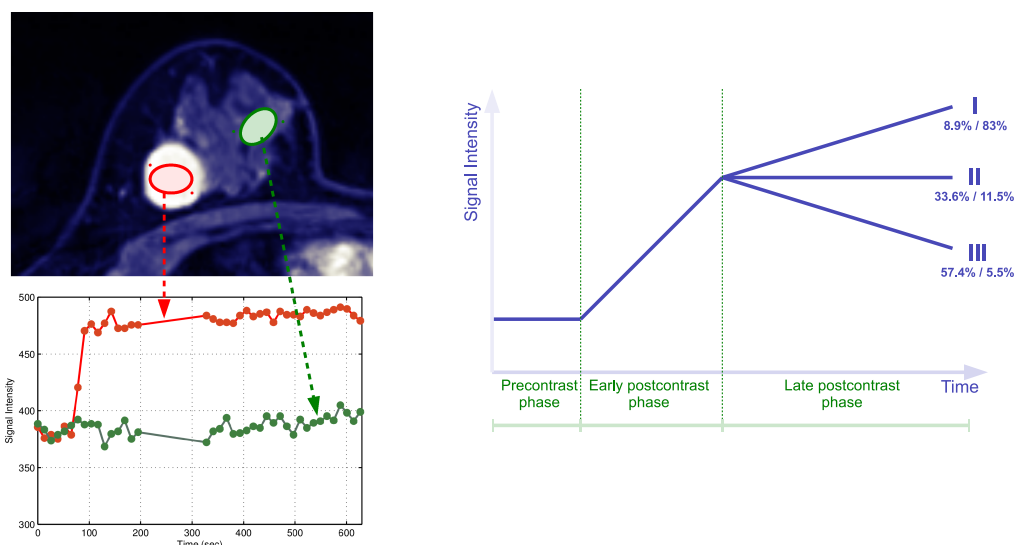
Figure 1.3: Left: DCE-MRI Time-Signal intensity curve extracted from a ROI within cancerous tissue (red) and normal tissue (green). Right: Schematic drawing of the time-signal intensity curve phases and types (I - III). The percent numbers next to the curve type represents the likelihood of a malign / benign lesion having the specific type of curve, as determined by Kuhl et al. [87]

II), or whether it declines again (Type III or washout curve) (Figure 1.3 right). Invasive cancers tend to have a curve with a steep uptake in combination with a washout (Type III) or plateau (Type II) [85], whereas benign lesions tend to have either a low contrast uptake rate or have a persistent increasing enhancement (Type I). A detailed description of characterizing lesions by interpreting the kinetic curves and the morphology (shape, margins...) is given in Kuhl et al. [85]. In Figure 1.3 a time-signal curve from a contrast-enhancing cancerous tissue region and non-contrast enhancing normal tissue region is presented. Furthermore, a schematic overview of the contrast enhancement phases and kinetic curve types is given.

A sensitivity of 89-100% is reported using the DCE-MRI modality. On the other hand, the reported specificity is rather variable, ranging from 37 to 97% [85]. In the last years PET and DWI have been proposed as additional modalities to increase the specificity of MRI [108, 171, 172].

1.5.2 Diffusion Weighted Magnetic Resonance Imaging (DWI)

DWI measures the diffusivity of water molecules in tissues, providing an insight into tumor cellularity [116]. A compact cellular structure, as occurs in malignant tumors, results in a reduced extracellular spacing and limited water diffusion. In contrast, in benign lesions, where the extracellular space is larger, the diffusion of water is less restricted [116].

The main parameter in the DWI acquisition process is the diffusion attenuation, denoted as *b-value* (sec/mm^2) [16]. It can be seen as a parameter for the sensitivity to diffusion. Images

6
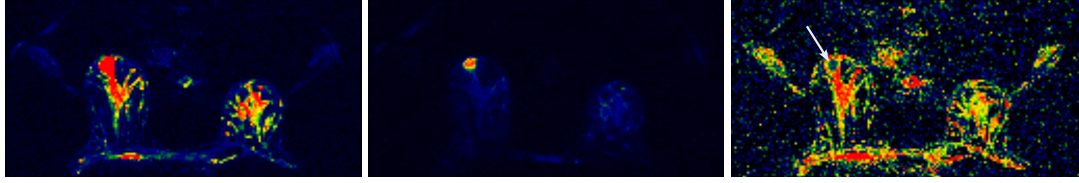
Figure 1.4: Diffusion Weighted Magnetic Resonance Images. Left: b value protocol of 0 sec/mm^2. Middle: b value protocol of 850 sec/mm^2. An identical scaling and color coding is used for both images. Blue color refers to low DWI signal value, red color refers to a high DWI signal value. Right: ADC mapping. The arrow point out the malign lesion. A reduced ADC value is observed within the lesion, in comparison to the surrounding tissue.

acquired with a low b-value are less diffusion weighted than images with a high b-value. On low b-valued images more morphologic structures are visible. Images acquired with high b-values have a lower Signal-to-Noise Ratio (SNR) and more image distortion. On the other hand, on high b-valued images malignant tumors are highlighted, due to the restriction of the water diffusion. This effect is illustrated in Figure 1.4.

The signal intensity values in DWI are influenced by MRI T_1 and T_2 relaxations in addition to diffusion amount [16]. In order to reduce the relaxation effect in DWI images and to get more "pure" diffusion values, an Apparent Diffusion Coefficient (ADC) map is calculated from two (or more) measurements with different b-values, b_1 and b_2, by using following formula [16]:

$$ADC(x,y,z) = \ln\left[\frac{S_2(x,y,z)}{S_1(x,y,z)}\right] / (b_1 - b_2) \tag{1.1}$$

S_1 and S_2 are the diffusion signal values from two images acquired with the different b-values.

The ADC value is correlated with the amount of diffusion and therefore inversely correlated with the compactness of the cellular structure of tumors. Malignant breast tumors exhibit a higher cellularity, resulting in a lower ADC value than for benign breast tumors (Figure 1.4 right). A simple threshold level of the ADC value may therefore be used to get a differentiation between benign and malign lesions. Several studies demonstrated the effectiveness of such a method [24, 65, 97, 171–173]. Guo et al. [65] reports a sensitivity and specificity of 93 % and 88%, respectively. Marini et al. [97] reports a sensitivity of 80% and a specificity of 81%. Bogner et al. [24] evaluated the influence of the b-value parameter on the ADC-values and determined the optimal b-values. They report a sensitivity, respectively specificity of 96% and 94%.

1.5.3 Positron Emission Tomography (PET)

PET is a nuclear medicine imaging method, using a positron-emitting radionuclide (*tracer*). When using FluoroDesoxyGlucose (FDG) as tracer, a 3-dimensional functional image can be obtained in a PET scanner, describing the glucose metabolic activity on a molecular level [44]. The FDG tracer is taken up and accumulated in glucose consuming cells, resulting in an increased amount of positrons emitted at these regions. The annihilation of a positron with an
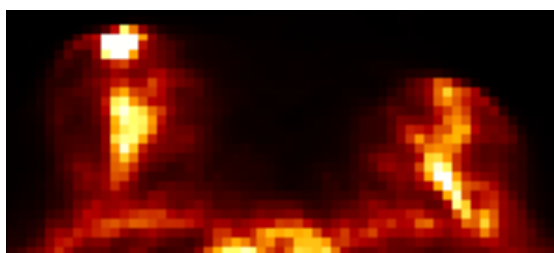
Figure 1.5: Breast PET image. The highlighted region in the top-left part of the image is a malignant lesion.

electron produces two gamma photons moving in opposite directions. These paired gamma photons are detected in the PET-scanner. By accumulating tens of thousands coincident events of gamma photons detected in the scanner, a 3-dimensional image can be reconstructed [44].

Liver, kidney, and brain tissues, but also many cancerous tissues have a high glucose metabolism, resulting in a higher signal intensity in the functional PET images (Figure 1.5). Therefore, the metabolic information provided by the PET scanner is useful in lesion localization, in differentiating benign and malignant breast lesions, in disease staging, and assessment of treatment response [180].

However, PET suffers from a very low resolution, which is 4 - 6 mm for commercial scanners [38], a low SNR, and lack of morphologic information. This renders the localization of lesions difficult, and prevents the detection of small lesions. To overcome this limitation hybrid PET / CT scanner [147] and recently PET / MRI scanner [5, 78] have been proposed, where the functional PET image is fused with a morphologic CT or MR image.

The potential of fused PET-MRI in breast cancer diagnosis have been shown in several studies [108, 117, 120]. Moy et al. showed that fusing breast images of PET and MRI increases the specificity of MRI, whereas in this study only the morphologic information of MRI is used. Initial results of a study by Pinker et al. [117, 120], where PET, DCE-MRI, and DWI is combined, demonstrated promising results, with a reported sensitivity of 98 % and a specificity of 78 % in the diagnosis of breast cancer and lymph-node metastasis. This master thesis is partially founded on and motivated by the study of Pinker et al.

Multimodal Medical Image Registration

This chapter describes the multimodal registration of breast images. The aim of the registration process, as used in this thesis, is the transformation of the multimodal dataset into one unified coordinate system, so that an accurate spatial relationship of the images is ensured. In this application the relationship is the anatomical correspondence. In other applications functional correspondences (e.g. lining up functionally equivalent regions of the brains) or structural–functional correspondences (e.g. correctly positioning functional information on a structural image) [43] are used for registration.

As Guo et al. [66] stated in their survey: *"the challenging task of breast-image registration is the inhomogeneous, anisotropic nature of the soft-tissue within the breast, its nonrigid body behavior and the various imaging conditions."*. Spatial variations in the multimodal breast images are inherent in the image acquisition process: i.e. breathing and slight patient movement during a DCE-MRI acquisition; varying patient positions in the MRI and the PET/CT scanner; anisotropic image distortions of the DWI modality [74]. Figure 2.1 illustrates such spatial variations. Advanced nonrigid registration methods are necessary, which are able to handle *large deformations* as well as *intermodal* registration.

The aim of this chapter is to give an overview of the state-of-the-art image registration methods, with the focus on registration of multimodal breast images. In the following sections the essential components of a registration framework are discussed, covering the topics of image similarity measures in Section 2.2, transformation models (Section 2.3) and an outline of optimization strategies in Section 2.4.

2.1   Theoretical Overview of a Registration Framework

The registration process can be seen as an optimization process, where a spatial mapping from image $I(\mathbf{x})$ to image $J(\mathbf{z})$ is obtained in a way that both images are optimally aligned. This is

Figure 2.1: Spatial variations of the acquired images. Left: MR image (red) overlaid by a CT image (blue/white). A nonrigid displacement is observed between these two modalities, due to different scanning positions and orientations for MR and CT images. Middle: Overlay of two MR images acquired in a DCE-MRI time-series. A small shift between these two images is visible, as a result of small patient movement and/or breathing in the acquisition process. Right: MR image (red) overlaid by a DW image (white). The anisotropic image distortion of the DW image, which is inherent in this modality, causes the mismatch of these two images. All images are thresholded for illustration purpose.

done by finding a transformation $\mathbf{T}$ that maps all locations $\mathbf{x}$ from image $I(\mathbf{x})$ to image $J(\mathbf{z})$ by minimizing a cost function C:

$$\mathbf{T}_{optimal} = \arg\min_{\mathbf{T}} \sum_{\mathbf{x}} C\left(I(\mathbf{T}(\mathbf{x})), J(\mathbf{z})\right) \tag{2.1}$$

In this thesis image $I(\mathbf{x})$ is the input image that is going to be transformed, and $J(\mathbf{z})$ is called the *reference* image, to whom the input image is registered to. Formally an image I represents a data function $I : \Omega \to \mathbb{R}^d$, where $\Omega \subseteq \mathbb{R}^n$ is the image domain($n = 2$ is 2D domain, $n = 3$ is 3D domain) on which the data are defined. d defines the dimension of the data, i.e. $d = 1$ are scalar values, such as images acquired with MRI, $d = 3$ are color images or vector images. The transformation $\mathbf{T}$ can then be seen as a mapping in the image domain: $\mathbf{T} : \Omega \to \Omega$.

A registration framework is derived from the optimization problem, containing three main elements: The *transformation model*, the *cost function*, and the *optimization strategy*. The transformation model describes the kind of transformation applied on the registration, like *linear* transformations or *deformable* transformations. The cost function is realized as a *similarity metric* $\Pi_\sim$ quantifying the amount of similarity, often in combination with a regularization or penalization term Π_R. Since the optimization problem is ill-posed, regularization adds additional user knowledge to the problem, i.e. by applying physical models like the fluid or elastic model. Since finding the optimal solution is NP-complete for non-rigid registration [80], iterative optimization techniques are commonly used for finding the approximate optimal solution. The choice of the transformation model, the similarity metric, and the optimization strategy is highly depending on the application [30].

2.2 Similarity Metrics

The similarity metric measures how "similar" two images are, where similarity is defined by the amount of correspondences between two images. It is an essential part of the cost function that

is optimized. The choice of the similarity metric is highly depending on the application. Giving an general overview of all similarity metrics is far out of scope of this thesis. In the following part metrics suitable for medical image registrations are reviewed.

The similarity measures can be divided into two main categories, the geometric based, and the intensity based. In geometric approaches correspondences are defined on anatomical elements identifiable in both images. These elements can be point landmarks, curves, or surfaces. The correspondence points guides the registration, where points in between are interpolated [130]. Typically, the distances among physical points are minimized in the optimization process. The paired landmarks are either defined manually [130] or automatically by identifying anatomical feature correspondences, like nipple position, breast shape [166, 176], or geometrically by identifying points, edges, curves or surfaces using the voxel intensities and the gradient [20, 94, 149, 150]. Point landmarks may also be defined by physical objects attached to the scanning subject that are clearly visible in the image modalities, like skin fiducial markers [162].

Intensity based approaches do not use any anatomical knowledge, instead intensity patterns are matched over the whole image. A mathematical or statistical similarity metric is defined, which is used to adjust the transformation model until the similarity function is maximized (or the error is minimized). It is assumed, that the images are registered when they are most similar. Hermosillo et al. reviewed several intensity similarity metrics [71], where the three common metrics Mean Squared intensity difference (MSQ) [40, 148, 156], Cross-Correlation (CC) [12, 58], and Mutual Information (MI) [98, 132, 158] are reviewed in this thesis.

2.2.1 Mean Squared Intensity Difference (MSQ)

A simple and intuitive metric is the MSQ, also titled as Sum of Squared Differences (SSD), which assumes identical intensities in both images, varied only by Gaussian noise. This metric is therefore only suitable for intramodal image registrations. It is defined as:

$$MSQ = \frac{1}{N} \sum_{\mathbf{x}} \left(I(\mathbf{T}(\mathbf{x})) - J(\mathbf{x}) \right)^2 \tag{2.2}$$

An optimal match results in a zero valued MSQ. The advantage of this metric is that it is simple to compute [76] and it has a relatively large capture radius [177]. The main disadvantages are its restriction to intramodal image registration, and its sensitivity to small number of outlier voxels [43]. Also, a linear change in intensity values results in a poor similarity score [177].

2.2.2 Cross-Correlation (CC)

This metric, also denoted as *correlation coefficient* or *normalized correlation*, computes the pixel-wise cross-correlation of the two images, normalized by the square root of the autocorrelation of each image. A linear relationship of the image intensities is assumed, unlike the MSQ metric, where an identical intensity relationship is assumed. This makes the CC metric suitable for intramodal as well as for some intermodal registrations. Defining the variables $\overline{J}(\mathbf{x}) = J(\mathbf{x}) - \mu_{J(\mathbf{x})}$ and $\overline{I}(\mathbf{x}) = I(\mathbf{T}(\mathbf{x})) - \mu_{I(\mathbf{T}(\mathbf{x}))}$, where $\mu_{J(\mathbf{x})}$ is the mean intensity value

of the image $J(\mathbf{x})$, leads to following formulation of the metric:

$$CC = \frac{\sum_{\mathbf{x}} \overline{I} \cdot \overline{J}}{\sigma_I \cdot \sigma_J} \tag{2.3a}$$

$$= \frac{\sum_{\mathbf{x}} \overline{I} \cdot \overline{J}}{\sqrt{\sum_{\mathbf{x}} \overline{I}^2 \cdot \sum_{\mathbf{x}} \overline{J}^2}} \tag{2.3b}$$

$$= \frac{\langle \overline{I}, \overline{J} \rangle}{\|\overline{I}\| \cdot \|\overline{J}\|} \tag{2.3c}$$

$\|$ is the L^2-Norm and $\langle \cdot \rangle$ is the vector inner product. The optimum value of this metric is 1 when both images match perfect, whereas misalignments results in lower values.

CC is robust to variations in lighting and exposure conditions [92]. Avants et al. [12] proposed a windowed version of CC for medical image registration, where the coefficient is calculated from a sliding window. They showed that this version adapts well to situations with locally varying intensities, like MRI inhomogeneities (bias field).

2.2.3 Mutual Information (MI)

This metric has been introduced independently by Viola and Wells [158] and by Maes et al. [96] as a information theoretic measure. Unlike the CC metric, where a functional intensity relationship is assumed, MI assumes only a probabilistic relationship among the voxel intensities. The image intensities of the two images are considered as two random variables A and B. MI measures how much information of the second random variable is in the first random variable by using the *Shannon entropy* [138]. The entropy variables are defined as following:

$$H(A) = -\sum_a p_A(a) \cdot \log p_A(a) \tag{2.4a}$$

$$H(B) = -\sum_b p_B(b) \cdot \log p_B(b) \tag{2.4b}$$

$$H(A, B) = -\sum_{a,b} p_{AB}(a, b) \cdot \log p_{AB}(a, b) \tag{2.4c}$$

where $H(A)$ respectively $H(B)$ are the entropies of the two random variables, and $H(A, B)$ is the joint entropy. p_A is the marginal probability density function for random variable A, and p_{AB} is the joint probability density function. If A and B are completely unrelated, then the joint entropy is the sum of the individual entropies:

$$p_{AB}(a, b) = p_A(a) \cdot p_B(b) \tag{2.5a}$$
$$H(A, B) = H(A) + H(B) \tag{2.5b}$$

If they are related, then $H(A, B)$ is smaller than $H(A) + H(B)$. The difference between the joint entropy and the marginal entropies is called Mutual Information:

$$MI(A, B) = H(A) + H(B) - H(A, B) \tag{2.6}$$

MI is zero if the two random variables are independent, whereas with an increasing dependency of the random variables MI increases too.

The advantage of using this probabilistic approach is that no specific form of dependency is necessary [177], in contrast to the two previously reviewed metrics, where an intensity similarity, respectively a linear dependency is assumed. This makes MI suitable for multi-modality registration [122]. MI, applied on a global scale, is known as a useful metric for a robust rigid registration [12]. However, the performance of MI on a local scale, as it is necessary for deformable registration, is reduced when non-stationary noise patterns or intensity inhomogeneities in the image (i.e. MRI bias fields) exists [12, 143]. The joint-probabilities have to be calculated from a large number of samples to get meaningful statistics, resulting in a trade-off between locality of the MI estimation and the statistical reliability [12].

Joint Histogram   In an image registration problem the marginal and joint probability densities are in common not available and have to be estimated from the image intensities. This can be done by using either a *joint histogram* for discrete values or *Parzen windowing* [114, 131] for continuous values. As an alternative, continuous values can be discretized by binning them. Parzen windows are also used to smooth the discrete histogram probability densities yielded from the joint histogram [98].

The joint histogram $h(a, b)$ is a 2-dimensional $n \times m$ matrix, where n and m are the number of (binned) discrete intensity values in the images I and J. Entry (a, b) of the matrix contains the number of times intensity a in the first image and b in the second image appear at the same spatial location. The probability densities are estimated from the normalized joint histogram: $\hat{P}_{AB}(a, b) = h(a, b)/N$, where N is the sum over the histogram. The marginal probability densities are estimated by summing over the rows, respectively columns, of the normalized histogram: $\hat{P}_A(a) = \sum_{b=1}^{m} \hat{P}_{AB}(a, b)$, and $\hat{P}_B(b) = \sum_{a=1}^{n} \hat{P}_{AB}(a, b)$.

Parzen Windowing   Parzen windowing, also known as *kernel density estimators*, is a non-parametric method of estimating the probability density function of a random variable. Intensity samples S are taken at random positions from the image and a kernel function $K(\cdot)$ is placed on the samples. The estimated probability density is the superposition of all kernels:

$$\hat{P}(a) = \frac{1}{nh} \sum_{s_j \in S} K\left(\frac{a - s_j}{h}\right) \tag{2.7}$$

n is the number of samples in the set S, and $h > 0$ is the smoothing parameter or bandwidth, which is used to scale the kernel. The kernel function needs to be smooth, symmetric, have zero mean and integrate to one, like Gaussian, boxcar, or B-spline functions. The choice of the smoothing parameter h is always a trade-off between the bias of the estimator and its variance. A large h may smooth out features, like the modes of the density, whereas too small h may result in a too noisy density distribution. Figure 2.2 illustrates an estimation using the Gaussian as kernel function.

To get discrete values from the continuous density estimation, which are needed in the entropy calculation (Equation 2.4) two methods have been proposed: In the method of Viola and
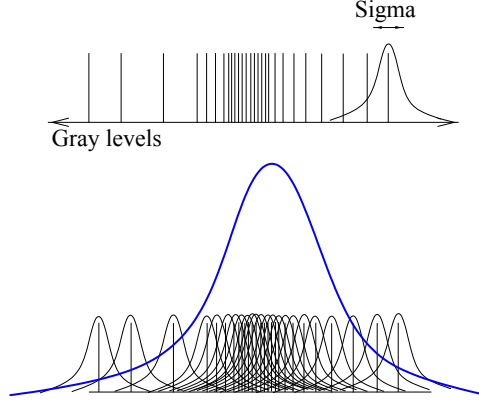
Figure 2.2: Illustration of the Parzen windowing. A kernel function (here Gaussian) is placed at each sample point (top). A continuous density function is constructed by superimposition the kernel functions (bottom). Image taken from [76].

Wells [158] a second sample set R is drawn from the image, in addition to the probability distribution estimating set S. The entropy is then approximated as a sample mean:

$$H(A) = \frac{1}{N} \sum_{r_j \in R} log \hat{P}(r_j) \tag{2.8}$$

Mattes et al. [98] uses only one sample set S to estimate the probability densities using B-Spline kernel functions. The entropy is then approximated by evaluating the probability densities at equally spaced discrete positions or bins.

Normalized Mutual Information Studholme et al. [144] proposed a normalized version of the MI metric, dealing with the problem that the MI metric may increase with *increasing* misregistration for cases of a small initial overlap. If the relative areas of background and object are in balance, the sum of the marginal entropies may increase faster than the joint entropy [122]. The normalized mutual information metric is:

$$NMI(A, B) = 1 + \frac{MI(A, B)}{H(A, B)} = \frac{H(A) + H(B)}{H(A, B)} \tag{2.9}$$

where $1 \leq NMI \leq 2$.

Another variant of a normalized MI metric has been proposed by Maes et al. [96], namely the *Entropy Correlation Coefficient (ECC)*:

$$ECC(A, B) = 2 \cdot \frac{MI(A, B)}{H(A) + H(B)} \tag{2.10}$$

where $ECC = 2 - 2 \cdot NMI$ and $0 \leq ECC \leq 1$.

2.3   Transformation Model

The transformation model describes the spatial mapping of the input image coordinate system to the reference image coordinate system. The choice of the model defines the type and the number of possible deformations [43], and is dependent on the application. The transformations can be divided into following categories: [7]

- *Rigid - Non-Rigid*

- *Parametric - Non-parametric*

- *Global - Local*

In a rigid transformation the distances and proportions of the objects are preserved, whereas in a non-rigid transformation this is not the case in common. A rigid transformation is a combination of rotation and translation. Non-rigid transformations may be *affine* transformations, allowing shearing and scaling of the image; or an *elastic* transformation, allowing local deformations in the moving image.

Parametric transformations are defined by a limited set of parameters $\mathbf{p} = (p_1, ..., p_n)$. For instance, a two-dimensional rigid transformation is parameterized by a rotation angle θ and a translation vector $\mathbf{t}$. The number of parameters defines the *degrees of freedom* of a transformation. Non-parametric transformations usually operate on vector fields describing the displacement of each voxel [155].

Parametric transformations are divided into global and local transformations, according to the range of influence of a parameter on the transformation. Global transformations are for instance rigid and affine transformation, where a change of a parameter has a global influence. Another global transformation is the Thin-Plate Spline (TPS) transformation [27], where the transformation parameters are the positions of control points defining the shape of thin-plate-splines. Shifting a control point position affects all other points [43]. Global transformations have the disadvantage that the ability of modeling local deformations is limited [43].

In the subsequent section following transformations and their concepts are described:

- Rigid and Affine transformations,

- Spline-Based transformations,

- Demon-Based transformations, and

- Symmetric Normalization (SyN).

Rigid and affine transformations are used to match images on a global scale. Spline-Based transformations are able to register images containing non-rigid deformations, with a limited use of modeling large deformations, due to the small deformation constraint formulation [11]. Demon-based algorithms are known to be a fast and efficient registration method in a large deformation framework [156]. SyN is a large deformation diffeomorphic mapping. Such models are known for being able to generate a dense, smooth and invertible mapping (diffeomorphic) along with the capability of handing large deformations. SyN also supports various metrics

(MI,CC,MSQ, labeled point sets) [14]. These facts make this transformation of special interest for this thesis, where large non-rigid deformations in the multi-modality breast images have to be taken into account.

2.3.1 Rigid and Affine Transformation

A rigid transformation is a mapping of a Euclidean space to itself, where the distance between the points is preserved. A rigid transformed object has the same shape and size as the original object. Two geometrical operations and their combinations fulfill this definition: *rotation* and *translation*. A rigid transformation T applied on the coordinates $\mathbf{x}$ can be written as:

$$T(\mathbf{x}) = \mathbf{R}\mathbf{x} + \mathbf{t} \tag{2.11}$$

where R needs to be an orthogonal transformation ($R^T = R^{-1}$) defining the rotation, and $\mathbf{t}$ is the translation vector.

A superset of the rigid transformation is the affine one, adding the geometrical operations *scaling* and *shearing* to the transformation. The affine transformation can be written as a compact matrix $\mathbf{T}$, including the translation, by using homogeneous coordinates [167]. In the two-dimensional case the affine transformation has the form [61]:

$$\begin{bmatrix} x & y & 1 \end{bmatrix} = \begin{bmatrix} v & w & 1 \end{bmatrix} \mathbf{T} = \begin{bmatrix} v & w & 1 \end{bmatrix} \begin{bmatrix} t_{11} & t_{12} & 0 \\ t_{21} & t_{22} & 0 \\ t_{31} & t_{32} & 1 \end{bmatrix} \tag{2.12}$$

Depending on the parameters t_{11} to t_{32} a rotation, translation, shear, and/or scale of the co-ordinates are applied. Table 2.1 lists how the parameters influence the transformation. Affine transformations can be combined into a single affine matrix $\mathbf{T}$ by applying a matrix multiplication on them. This follows from the associativity of the matrix multiplication:

$$\mathbf{A}(\mathbf{B}\mathbf{x}) = (\mathbf{A}\mathbf{B})\mathbf{x} \tag{2.13}$$

It has to be noted, that the order of the multiplication is important, since the matrix multiplication is in common not commutative: $\mathbf{A}\mathbf{B} \neq \mathbf{B}\mathbf{A}$.

The affine and rigid transformations are global and parametric transformations. Local deformations cannot be modeled by them. This limits the usage of affine transformations for breast image registration, since the breast is highly non-rigid, due to the soft-tissues in the breast. However, rigid transformations can be used as initial registration to minimize the global alignment error, followed by a deformable transformation modeling the local deformations [14, 132].

2.3.2 Spline-Based Transformations

This kind of transformation relies on spline functions and the spatial variation of their control points to model a non-rigid transformation. They are suitable for point-based landmark registrations [62] as well as for voxel registration using an intensity-based metric [132].

Goshtasby [62] proposed a point-based landmark registration based on TPS for registration. The corresponding landmark points are used as control points for the splines. The landmarks

Transformation Name	Parameter	Affine Matrix T	Coordinate Equations
Identity	none	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x = v$ $y = w$
Scaling	c_x, c_y	$\begin{bmatrix} c_x & 0 & 0 \\ 0 & c_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x = c_x v$ $y = c_x w$
Rotation	θ	$\begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x = v\cos\theta - w\sin\theta$ $y = v\sin\theta + w\cos\theta$
Translation	t_x, t_y	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ t_x & t_y & 1 \end{bmatrix}$	$x = v + t_x$ $y = w + t_y$
Shear (vertical)	s_v	$\begin{bmatrix} 1 & 0 & 0 \\ s_v & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x = v + s_v w$ $y = w$
Shear (horizontal)	s_h	$\begin{bmatrix} 1 & s_h & 0 \\ s_v & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x = v$ $y = s_h v + w$

Table 2.1: Two-dimensional affine transformations: $(x, y) = \mathbf{T}[(v, w)]$.

are displaced in the first image, so that they match with the landmarks in the second image. The deformations in between the landmarks are interpolated using TPS. Figure 2.3 illustrates the TPS-based transformation. Since the influence of the control points on the spline-shape is global, a movement of one control point affects all other points in the image. This global behavior has two side-effects: adapting the model to match local deformations is limited, and the computational costs when moving a single point rises steeply with increasing number of control points [43].
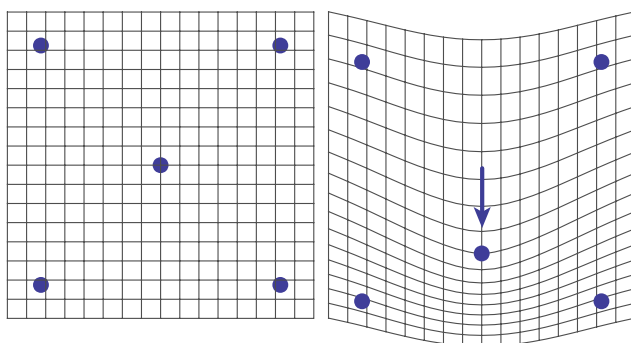
Figure 2.3: Illustration of a TPS-based transformation. The blue circles are the corresponding landmarks from a reference image (left) and a moving image (right). The TPS function is applied on a regular grid to illustrate the deformation of the moving image when moving the corresponding center landmark to the matching position. Image adapted from [105].
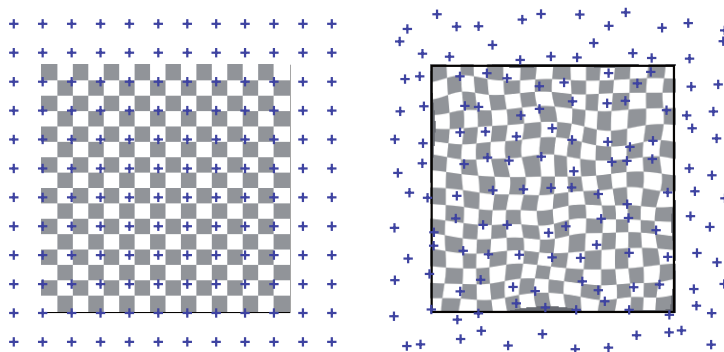
Figure 2.4: Illustration of a FFD-transformation. Initial configuration (left) and deformation by displacing the control points (right). Image adapted from [135]

.

To overcome these problems Rueckert et al. [132] suggested in his milestone paper a Free-Form Deformation (FFD) model [136] using B-Splines [89]. A FFD is based on a mesh-grid of control points, where the deformation is modeled by displacing these grid points. Figure 2.4 illustrates such a deformation. Using B-Splines have the advantage that they are locally controlled, in contrast to TPS, making them computational efficient even for a large number of control points, and a change of a control point influences only the local neighborhood.

The deformation transformation works as follows: First a mesh-grid Φ of control points $\Phi_{i,j,k}$ with uniform spacing δ is created. The FFD transformation can be written as a 3-D tensor product of 1-D cubic B-Splines:

$$\mathbf{T}(\mathbf{x}) = \sum_{l=0}^{3} \sum_{m=0}^{3} \sum_{n=0}^{3} B_l(u) B_m(v) B_n(w) \Phi_{i+l,j+m,k+n} \qquad (2.14)$$

B_l is the lth basis function of the B-Spline:

$$B_0(u) = (1 - u)^3/6$$
$$B_1(u) = (3u^3 - 6u^2 + 4)/6$$
$$B_2(u) = (-3u^3 + 3u^2 + 3u + 1)/6$$
$$B_3(u) = u^3/6$$

In an optimization process the positions of the control points Φ are varied until the cost function is minimized. The cost function is consisting of a similarity metric term $\Pi_\sim$ and a penalty term Π_R as regularization:

$$C(\Phi) = \Pi_\sim(J(\mathbf{x}), I(\mathbf{T}(\mathbf{x})) + \lambda \Pi_R(\mathbf{T}) \tag{2.15}$$

The penalty term forces the FFD transformation to be smooth, where the λ is a weighting parameter of the penalty term, influencing the smoothness of the transformation. In 3-D, the penalty term is defined as following:

$$\begin{aligned}
\Pi_R = \frac{1}{V} \int_0^X \int_0^Y \int_0^Z &\Big[\Big(\frac{\delta^2 \mathbf{T}}{\delta x^2}\Big)^2 + \Big(\frac{\delta^2 \mathbf{T}}{\delta y^2}\Big)^2 + \Big(\frac{\delta^2 \mathbf{T}}{\delta z^2}\Big)^2 \\
&+ \Big(\frac{\delta^2 \mathbf{T}}{\delta xy}\Big)^2 + \Big(\frac{\delta^2 \mathbf{T}}{\delta xz}\Big)^2 + \Big(\frac{\delta^2 \mathbf{T}}{\delta yz}\Big)^2 \Big] dx \, dy \, dz
\end{aligned} \tag{2.16}$$

where V is the volume of the image. The second-order mixed partial derivatives used in the penalty term are a measure of the smoothness [160].

The control points Φ are the parameters of the B-Spline FFD. Few control points with a large spacing are only able to model global nonrigid deformations, whereas a fine mesh of control points is able to model local deformations. A higher degree of freedom increases also the computational effort. Therefore a trade-off between model flexibility and computational complexity is given [132]. Rueckert et al. proposed a hierarchical refinement approach, starting with an initial large spacing, and successively halving the spacing between the control points by adding new ones using a B-Spline subdivision algorithm [54].

2.3.3 Demons Based Transformation

In 1998 Thirion proposed a non-rigid and non-parametric image registration method treating the non-rigid registration as a diffusion process [148]. He named it "Demons Algorithm", where the "demons" term is referring to a concept introduced by Maxwell in the 19th century, solving the Gibbs paradox in the thermodynamics.

In Thirions approach demons are forces that push according to the optical flow equation [17]. For small displacements of a given point $\mathbf{x}$ the optical flow equation is:

$$\mathbf{u} \cdot \nabla I(\mathbf{x}) = I(\mathbf{x}) - J(\mathbf{x}) \tag{2.17}$$

where $\mathbf{u}$ is the displacement, and $\nabla I(\mathbf{x})$ is the gradient of image I at the spatial point $\mathbf{x}$. From this equation Thirion derived the Demons formulation:

$$\mathbf{u} = \frac{(I(\mathbf{x}) - J(\mathbf{x})) \nabla I(\mathbf{x})}{|\nabla I(\mathbf{x})|^2 + (I(\mathbf{x}) - J(\mathbf{x}))^2} \tag{2.18}$$

In this case the displacement is guided by two forces, the internal edge based force $\nabla I(\mathbf{x})$ and the external force $I(\mathbf{x}) - J(\mathbf{x})$. The term $(I(\mathbf{x}) - J(\mathbf{x}))^2$ has been added by Thirion to make the equation more stable. Wang et al. [161] enhanced this equation by considering also the image edge forces $\nabla J(\mathbf{x})$ of the second image, and Cachier et al. [32] added a normalization term α to the equation, leading to following formulation:

$$\mathbf{u} = \frac{(I(\mathbf{x}) - J(\mathbf{x}))\nabla I(\mathbf{x})}{|\nabla I(\mathbf{x})|^2 + \alpha^2(I(\mathbf{x}) - J(\mathbf{x}))^2} + \frac{(I(\mathbf{x}) - J(\mathbf{x}))\nabla J(\mathbf{x})}{|\nabla J(\mathbf{x})|^2 + \alpha^2(I(\mathbf{x}) - J(\mathbf{x}))^2} \tag{2.19}$$

This equation is used in an iterative algorithm as an update rule for a displacement field $\mathbf{U}$. The displacement field is alternatively updated by the equation (2.19) and regularized by smoothing the displacement field with a Kernel K: $\mathbf{U} \leftarrow K * \mathbf{U}$, where K is typically a Gaussian Kernel, and $*$ is the convolution operator. The transformation $\mathbf{T}(\mathbf{x})$ is then:

$$\mathbf{T}(\mathbf{x}) = \mathbf{x} + \mathbf{U}(\mathbf{x}) \tag{2.20}$$

The original demons formulation as proposed by Thirion is limited to intramodal registration only, since it relies on image intensity differences. Therefore several enhancements to the Demons algorithm have been proposed, where either the intensities of one modality are transformed to the other modality [64, 84]; or by using an alternative formulation of the Demons registration introduced by Vercauteren et al. [154], where the intensity difference is replaced by other metrics, like normalized mutual information [106] or pointwise mutual information [95].

Vercauteren et al. proposed also a formulation of the Demons registration restricting the transformation to the group of diffeomorphisms [154–156]. They denoted it as Diffeomorphic Demons (DD). A detailed discussion of diffeomorphisms and its use in registration frameworks is given in the next section.

2.3.4 Symmetric Normalization (SyN)

SyN [14] is a state-of-the art large deformation, non-rigid, non-parametric transformation model, which operates in the group of diffeomorphisms, formulating the mapping as a *spatio-temporal* optimization problem [12].

A *diffeomorphic mapping* ϕ is an *invertible* and in both directions *differentiable bijective* mapping [50] that is closed under composition [14]. Restricting the mapping to the group of diffeomorphisms has the advantage that the topology of the anatomy is preserved and foldings that are physically impossible are prevented [156]. The mapping is always smooth, invertible and continuous parameterizable (i.e. in time) [177]. Invertibility is for instance a requirement when registering Diffusion Tensor Imaging (DTI) [175]. The diffeomorphic properties are also of interest in Computational Anatomy (CA), where the statistical variability of anatomical structures is analyzed, as in [8, 12, 90, 103, 104]. The diffeomorphic large deformation setting allows more shape variability to be modeled than in a small deformation environment, like B-Splines or elastic mapping [57] due to the fact that large deformation models penalizes deformation linearly, whereas small deformation models penalizes deformation quadratically [14].

The diffeomorphic approach was introduced by Trouve et al. [151] and has been enhanced by Miller et al. [103] by formulating the problem as an Euler-Lagrange equation. Based on

this work Beg et al. developed a large deformation formulation in the group of diffeomorphism, donating it as Large Deformation Diffeomorphic Metric Mapping (LDDMM) [19]. In this thesis a symmetric approach of the LDDMM is discussed, introduced by Avants et al. [12–14], namely SyN. This method is also part of the proposed CAD pipeline. Other proposals of diffeomorphic registration frameworks may be found in [9, 72, 155].

SyN operates in the diffeomorphic space $Diff_0$ by assuming a mapping of the image domain Ω to itself, and a homogeneous boundary condition such that the image border maps to itself: $\phi(\delta\Omega) = \mathbf{Id}$. Therefore rigid and scaling transformations have to be applied before the diffeomorphic mapping. The mapping ϕ, parameterized by the time parameter $t \in [0, 1]$, the spatial coordinate $\mathbf{x}$ and a smooth time-dependent velocity vector field $\mathbf{v}(\mathbf{x}, t) : \Omega \times t \to \mathbb{R}^d$, leads to a family of diffeomorphisms, $\phi(\mathbf{x}, t) : \Omega \times t \to \Omega$. These diffeomorphisms can be calculated as solutions of the ordinary differential equation (o.d.e):

$$\frac{d\phi(\mathbf{x}, t)}{dt} = \mathbf{v}(\phi(\mathbf{x}, t), t) \tag{2.21}$$

with respect to $\phi(\mathbf{x}, 0) = \mathbf{x}$. The spatial transformation ϕ is gained by integrating the velocity field $\mathbf{v}$ in time:

$$\phi(\mathbf{x}, 1) = \phi(\mathbf{x}, 0) + \int_0^1 \mathbf{v}(\phi(\mathbf{x}, t), t) dt \tag{2.22}$$

The existence and uniqueness theorem for o.d.e.'s [41], assuming a smooth vector field $\mathbf{v}$, ensures that the integration of Equation (2.21) generates a diffeomorphism [19]. From such a diffeomorphism a displacement or deformation field $\mathbf{U}$ is calculated in the following way

$$\mathbf{U}(\mathbf{x}) = \phi(\mathbf{x}, 1) - \mathbf{x}. \tag{2.23}$$

The time-parameterized transformation $\phi(\mathbf{x}, t)$ can be seen as a path in the space of diffeomorphic transformations, describing the movement of a particle in the velocity field over time, starting at position $\mathbf{x}$ at time $t = 0$ and ending at time $t = 1$ at the position $\phi(\mathbf{x}, 1) = \mathbf{z}$. The length s of such a path in the diffeomorphic space is calculated in a similar way than the arc length of a curve connecting two points in the Euclidean space:

$$D(\phi(\mathbf{x}, 0), \phi(\mathbf{x}, 1)) = \int_0^1 \|\mathbf{v}(\mathbf{x}, t)\|_L \, dt \tag{2.24}$$

where $\|\cdot\|_L$ is a Sobolev Norm and L is a linear operator. More details about the linear operator L are given below. A *geodesic* path, which is per definition the shortest path of two points in a space (here diffeomorphic space), is found by taking the *infimum* of all paths formed by Equation 2.24. The length of such a path gives a metric distance measure [19]. In an optimization process such a geodesic path between two diffeomorphism transformations $\phi(\mathbf{x}, 0)$ and $\phi(\mathbf{x}, 1)$ is going to be found by minimizing the distance between them. $\phi(\mathbf{x}, 0)$ is usually set to the identity transformation $\mathbf{Id}$. Combining a similarity metric $\Pi_\sim$ with the geodesic metric, which may be alternatively seen as a diffeomorphic regularization Π_R, leads to the LDDMM variational optimization problem introduced by Beg et al. [19]:

$$\mathbf{v}^* = \arg\min_{\mathbf{v}} \left\{ \int_0^1 \|L\mathbf{v}\|^2 dt + \lambda \int_\Omega \Pi_\sim(I, \phi(\mathbf{x}, 1), J) d\Omega \right\} \tag{2.25}$$
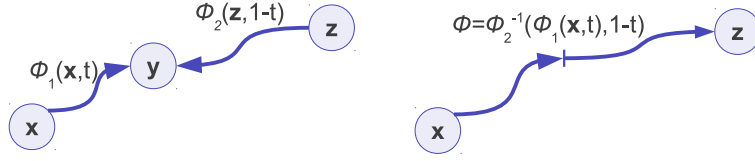
Figure 2.5: Left: An illustration of traversing a geodesic path, in time, from two end-points **x** and **z** to a midpoint **y** via ϕ_1 and ϕ_2. Right: Traversing from one end-point to the the other end-point by inverting the second diffeomorphism ϕ_2^{-1} and combining it with the first diffeomorphism ϕ_1. Image adapted from [13].

where λ controls the weight of the similarity measure. L is a linear differential operator and regularizes the velocity field $\mathbf{v}$, to ensure that the velocity field remains smooth in the optimization process. Smoothness of the velocity field is a prerequisite of calculating the diffeomorphisms. As linear operator a fluid regularization, like $\nabla^2 + \mathbf{Id}$, Navier-Stokes, or Cauchy-Navier operator is suggested [19,40]. The linear operator is either introduced as a term in a variational minimization [19] or in terms of a convolution with a Green's kernel function K: $\mathbf{v}_{smooth} = K * \mathbf{v}$. As kernel function K a Gaussian with varying σ may be used [14, 34, 156], where σ controls the "smoothness" of the velocity field. A detailed discussion about velocity field regularization is given in Dupuis, Grenader and Miller [49].

Avants et al. constructed a symmetric alternative of the LDDMM Equation (2.25) [13], considering the fact that a diffeomorphism ϕ is closed under composition, and may be decomposed into two components ϕ_1 and ϕ_2, as well as the fact that a geodesic path is symmetric and points on the path $\phi(\mathbf{x}, t)$ can be parameterized from both endpoints: $\phi_2(\mathbf{z}, 1 - t) = \mathbf{y} = \phi_1(\mathbf{x}, t)$. A full path ϕ and its inverse ϕ^{-1} can therefore be found by combining the partial paths ϕ_1 and ϕ_2 using the composition: $\phi(\mathbf{x}, 1) = \phi_2^{-1}(\phi_1(\mathbf{x}, t), 1 - t)$ and $\phi^{-1}(\mathbf{z}, 1) = \phi_1^{-1}(\phi_2(\mathbf{z}, t), 1 - t)$. Figure 2.5 illustrates the parameterization from both endpoints and the composition of the two transformations ϕ_1 and ϕ_2 to ϕ. In contrast to the LDDMM formulation, where the transformation is found at the endpoint of a geodesic path, beginning from the coordinate system of a moving image I to the reference image J, the symmetric formulation develops the transformation from both endpoints to a fixed point midway of the geodesic path. The full transformation ϕ is gained then by composition, as described above. The fixed point may be anywhere in the interval $[0, 1]$, whereas setting it to $t = 0.5$ has the effect, that the transformation is equally distributed on both images I and J. The terms "moving" image and "reference" image become obsolet in this environment. This approach guarantees sub-pixel accurate, invertible transformations [12]. Theoretically all diffeomorphisms are guaranteed to be symmetric and invertible, in practice interpolation errors summing up linearly with the number of interpolation steps can cause invertibility errors. The symmetric solution minimizes this kind of error by exploiting the diffeomorphic invertibility [12]. Figure 2.6 illustrates the transformation of two shapes along a diffeomorphism to the midpoint of the geodesic path.

Defining $\mathbf{v}(\mathbf{x}, t) = \mathbf{v}_1(\mathbf{x}, t)$ in the interval $t \in [0, 0.5]$, and $\mathbf{v}(\mathbf{x}, t) = \mathbf{v}_2(\mathbf{x}, 1 - t)$ in the
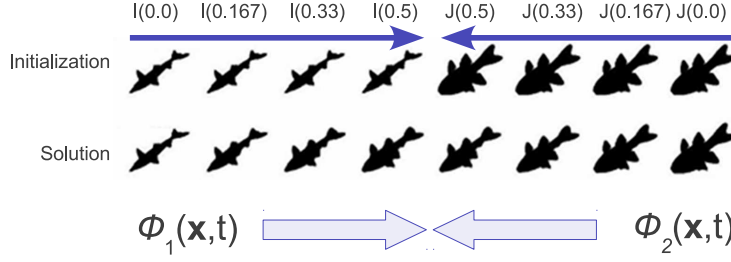
Figure 2.6: An illustration of the deformation of two shapes I and J along the geodesic paths ϕ_1 and ϕ_2 to a midpoint, forming a mean shape. The top row shows the original images, which are used as initialization of the SyN method. The bottom row shows the images after convergence of the SyN solution for the given time-point $t \in [0, 0.5]$. Image adapted from [12].

interval $t \in [0.5, 1]$ leads to the symmetric variant of Equation (2.25) [14]:

$$\{\mathbf{v}_1^*, \mathbf{v}_2^*\} = \arg\min_{\mathbf{v}1,2}\Big\{ \int_0^{0.5} \|L\mathbf{v}_1(\mathbf{x}, t)\|^2\, dt + \int_0^{0.5} \|L\mathbf{v}_2(\mathbf{x}, t)\|^2\, dt$$
$$+ \lambda \int_\Omega \Pi_\sim(I \circ \phi(\mathbf{x}, 0.5), J \circ \phi(\mathbf{x}, 0.5))d\Omega\Big\} \qquad (2.26)$$

A solution of the variational optimization problems (2.25) and (2.26) is found by using Euler-Lagrange equations [103]. A discussion of developing and solving such Euler-Lagrange equations is beyond the scope of this discussion. One may be referred to Beg et al. [19] and Avants et al. [12] for a detailed description.

Geodesic and Greedy SyN Avants et al. describes two variants of the numerical solution of Equation (2.26), the Symmetric Geodesic Normalization (SyGN) and the computational and memory lower-cost greedy variant, the Greedy Symmetric Normalization (GrSyN) [14]. The former is closer to the theoretical diffeomorphic formulation of the Equation by using a dense-in-time gradient calculation along the geodesic path in the interval $[0, 0.5]$ and a reintegration of the diffeomorphisms after each iteration. In the greedy optimization of Equation (2.26) the gradient is only calculated at the midpoint of the full diffeomorphism. Instead of reintegration of ϕ_1 and ϕ_2 they are updated from the previous iteration using following Equation:

$$\phi_i(\mathbf{x}, 0.5) = \phi_i(\mathbf{x}, 0.5) + (\delta K * \nabla\Pi_i(\mathbf{x}, 0.5))\sim\phi_i(\mathbf{x}, 0.5),\ i \in 1, 2 \qquad (2.27)$$

where δ is the step size of the gradient based optimization process.

2.4   Optimization Strategy

The optimization strategy defines how the transformation parameters, deformation fields or velocity fields are adjusted to improve the image similarity by minimizing the cost function C. Usually an iterative approach is used, where an initial parameter set is estimated, and iteratively

adjusted until a convergence criterion is achieved. At each iteration the cost function C is evaluated using the current parameter estimation. Then the parameters are adjusted in a way that the cost function is reduced. The optimization is terminated when some convergence criterion is achieved (i.e. cost function is not reducing anymore) [10]. Standard strategies are available for this kind of optimization, like Steepest Gradient Descent, Conjugate Gradient Method, Gauss-Newton Method, Powell's Method, Downhill Simplex Method and others [124]. Discussing these optimization methods is beyond the scope of this thesis. Readers interested in this topic may be referred to [26].

Gradient based optimization methods operate on the cost function derivative. Though, also the derivatives of the similarity functions are needed. In Hermosillo et al. [71] and Avants et al. [12] derivatives of the intensity similarity metrics are described.

Multiresolution Optimization The optimization solutions mentioned above are only able to find local minima of the cost function, which might be far away from the overall or global minimum of the cost function. To improve the robustness, accuracy and also the speed of the optimization process, a multi-resolution approach using a coarse-to-fine scheme is widely used [76]. The registration is first done on a coarse scale with a reduced amount of pixels or control points. On this level coarse deformations are identified, whereas small image variations are ignored. The spatial mapping determined at this level is then used as initialization of the registration at the next finer scale. These steps are continued until reaching the full scale. At the full scale level all image details are given, although the optimization process needs to handle just fine deformations, since larger deformations are already considered in the previous steps. This coarse-to-fine strategy improves the registration success rate and also increases robustness by eliminating local optima at coarser scales [76].

To get a coarse-to-fine scheme based on the image scale, a *Gaussian resolution pyramid* is used. At each level of the pyramid the resolution is halve of the resolution of the next level. Such a pyramid is recursively built by starting with the original image $I^{(0)}$ and recursively calculate the reduced resolution levels by applying a Gaussian smoothing to the image of the previous level, followed by a downscaling:

$$I^{(n+1)}(x, y, z) = (\omega \circledast I^n)(2x, 2y, 2z) \tag{2.28}$$

where ω is a discrete Gaussian smoothing kernel, $\circledast$ denotes a discrete convolution and n is the level index. The number of levels depends on the resolution of the original image and the amount of deformation. Avants et al. [14] suggests a 3 level pyramid for images with a resolution of 1 mm^3, and a sampling $n > 3$ in cases where the initial affine registration is weak.

2.5 Discussion

In this chapter the theoretical background of the registration process has been described with the focus on multimodal image registration. The aim of the registration process as used in this thesis is the transformation of the multimodal dataset into one unified coordinate system, so that an accurate anatomical relationship of the images is ensured. The registration is realized as an optimization process, where a transformation is iteratively developed and applied onto the

input image until the similarity with the *reference* image is optimized. The degree of similarity is measured by a metric. Three of these similarity metrics have been described in detail in this chapter, namely MSQ, CC, and MI. All of them are intensity based, where MSQ assumes identical intensities, CC assumes a linear relationship in the intensities, and MI assumes a probabilistic relationship of the input and reference image intensities. The transformation model is responsible for the spatial mapping of the input image to the reference image. The choice of the transformation model defines the degrees of freedom for the mapping. Four different types of transformation models have been described in this chapter: the rigid and affine transformations, suitable for global registration; spline-based transformations, where the image is deformed by varying control points of splines, and interpolating in between of them; Demon-based transformations, where a vector field is iteratively evolved, guided by the optical flow equation; and the LDDMM transformation, where the large deformation mapping operates in the space of diffeomorphisms, providing an invertible, topological preserving, dense and smooth mapping. SyN as a symmetric formulation of the LDDMM has been described in detail, where both images contribute to the registration process. Finally, the optimization process has been briefly described, where the parameters of the transformation model are iteratively modified until the cost function, which is a combination of the similarity metric and a regularization term, is minimized. A multiscale optimization based on a Gaussian resolution pyramid has been described, which is used to improve the robustness and the speed of the registration process.

3

Classification and Regression by Random Forests

Random Forests (RF) were developed and introduced by Leo Breiman and Adele Cutler [28]. It is an ensemble classifier, with the basic idea of growing many unpruned decision trees and using the statistical mode of the outcome of the trees to determine the class. Unpruned decision trees have a low bias, but suffer from a high variance, resulting in sensitivity to noise [69]. RF gains a low bias from the unpruned decision trees, and reduces the variance by aggregating many decision trees induced from bootstrap samples of the training data (*bootstrap aggregation*) and selecting random features during the tree building process [69]. An excellent overview of decision trees and RF is given by Hastie et al. [69]

RF is known to be an efficient algorithm with a comparable error rate to ADA-Boost [55], and robustness to noise [28]. The RF algorithm is used for classification or regression tasks. It is able to handle large feature sets without pre-selection, even if only a few variables carry the information necessary for the classification prediction, and the other variables are noise [28,133]. Additionally, the algorithm provides a measurement of the "importance" of features, which gives a more detailed insight into the contribution of features to the classification, and which has been shown to be useful for feature selection [6,59]. Another advantage of RF is that it has only three tuning parameter: *ntree*, the number of trees, *mTry*, the number of features considered at each split of a node, and *nodesize* the size of the leaves.

Writing a chapter about the RF algorithm is motivated by the fact that the RF algorithm is an integral part of this thesis. RF is used for binary classification in the segmentation part of the CAD pipeline to determine if a tissue is cancerous or non-cancerous, and in the classification part of the pipeline to distinct between benign and malign lesions. The variable importance measurement provided by the algorithm is used for feature analysis and feature selection in both cases.

The main topics of the chapter are the description of the RF algorithm in Section 3.1 and details about the measurement of the variable importance in Section 3.2.

3.1 The Random Forests Algorithm

The RF algorithm is a *supervised learning algorithm*. This means that the algorithm is trained using a training set $\mathcal{L}$ containing the input set $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\} \in \mathbb{R}^{n \times p}$ forming a p-dimensional feature space, and the corresponding output set $\mathbf{y} = \{y_1, y_2, ..., y_n\}$:

$$\mathcal{L} = \{(\mathbf{x}_i, y_i) | 1 \leq i \leq n\} \tag{3.1}$$

where n is the number of training examples. In case of classification the output $\mathbf{y}$ is of categorical type (i.e. class labels), whereas in the regression case the output is continuous. From the training set $\mathcal{L}$ a prediction model is generated. This model is used to predict an output value $\hat{y}$ from a new observation $\mathbf{x}_{new}$ according to the given model.

The RF prediction model is generated by following algorithm: Let n denote the number of training samples, p the number of feature variables (=dimension) of a sample, and T_i the ith tree in the forest.

For each tree T_b, $1 \leq b \leq ntree$:

1. Get a bootstrap sample $\mathbf{Z}^*$ from the training data by drawing m samples at random with replacement from the training data-set $\mathbf{X}$.

2. Recursively grow a tree T_i by applying following steps on each node:

   - Select $mTry$ feature variables at random from the p variables available

   - Find the best feature variable / splitting point among the $mTry$ variables and split the node into two sub-nodes

   - Repeat the recursive steps until the node size fall below a given minimum node size.

Breiman suggests as size for $mTry = \lfloor \sqrt{p} \rfloor$, and a minimum node size of 1 for classification. For regression, the default value is $mTry = \lfloor p/3 \rfloor$, and the minimum node size is five. As bootstrap size, m, he suggests to use about two-third of the instances. The unused samples are denoted as the Out-Of-Bag (OOB) samples [28]. These samples can be used to estimate the *generalization error* by predicting $\hat{y}$ for each sample x_i using only the trees of the Random Forest, in which the sample x_i is an OOB sample. The average of the errors made in the prediction is the OOB error rate, which has been shown to be an unbiased estimation of the *generalization error* [28]

Find the best feature variable / splitting point   In the algorithm a best split is applied at each node by determining the splitting variable and the splitting point giving the best binary partition of the node.

Regression trees   For regression trees one way of finding the best splitting point is the minimization of a sum-of-squares criterion, as proposed by Breiman et al. for their Classification And Regression Tree (CART) algorithm [29]. Denoting the splitting variable as j and the

splitting point (threshold) as s one may write the partition of the data in two regions R_1 and R_2 at a node as:

$$R_1(j,s) = \{\mathbf{X}|\mathbf{x}_j \leq s\} \text{ and } R_2(j,s) = \{\mathbf{X}|\mathbf{x}_j > s\} \tag{3.2}$$

Approximations for the splitting variable j and the splitting point s are found by solving:

$$\hat{j}, \hat{s} = \arg\min_{j,s} \left[\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - \hat{\mu}_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - \hat{\mu}_2)^2 \right] \tag{3.3}$$

$\hat{\mu}_1$ and $\hat{\mu}_2$ are determined for any choice of j and s by calculating the mean of y_i in the region R_1 and R_2

$$\hat{\mu}_1 = mean(y_i|\mathbf{x}_i \in R_1(j,s)) \text{ and } \hat{\mu}_2 = mean(y_i|\mathbf{x}_i \in R_2(j,s)) \tag{3.4}$$

This greedy algorithm tends to minimize the sum variances for the two resulting nodes.

Classification trees For classification trees the objective in splitting the node is that all samples in the sub-nodes should be ideally of the same class. The amount of "wrong" classes, or heterogeneity of target classes in a node is denoted as *impurity* of a node. The splitting variable and splitting point is chosen in a way that the biggest reduction of impurity in the sub-nodes is achieved.

Let $i(t)$ be a function measuring the impurity at node t. Furthermore let t_l and t_r be the left and right child node of node t, and n_t be the number of samples in node t. By partition the node according to a splitting variable j and splitting point s in a similar way as in Equation (3.2) one may calculate the probabilities P_l and P_r of the sub-nodes depending on the size of the two regions:

$$P_l(j,s) = P(x_j \leq s) = \frac{|R_1(j,s)|}{n_t} \text{ and } P_r(j,s) = P(x_j > s) = \frac{|R_2(j,s)|}{n_t} \tag{3.5}$$

where $|R_1(j,s)|$ is the number of samples in the left sub-node. The reduction of impurity ($\Delta i(t)$) between the parent node t and the child nodes t_l and t_r is then calculated as follows:

$$\Delta i(t) = i(t) - P_l \cdot i(t_l) - P_r \cdot i(t_r) \tag{3.6}$$

The best split is then determined by finding the biggest reduction of impurity by solving following maximization problem:

$$\hat{j}, \hat{s} = \arg\max_{j,s}[i(t) - P_l(j,s) \cdot i(t_l) - P_r(j,s) \cdot i(t_r)] \tag{3.7}$$

Measuring the node impurity   Two common methods of measuring the node impurity $i(t)$, producing almost similar results [31], are: *Gini index*, as used in the CART algorithm [29], and the concept of *information entropy*, as used in the *C4.5 decision tree* algorithm of Quinlan [125].

Let $p(k|t)$ be the conditional probability of class $k \in 1, ..., K$ at node t. An empirical determination of $p(k|t)$ is given by $\hat{p}(k|t) = n_k/n_t$, where $n_t t$ denotes the number of examples for node t and n_k the number of examples belonging to class k. The Gini impurity function $i_G(t)$ and the information entropy impurity function $i_E(t)$ are defined as following:

$$i_G(t) = \sum_{l \neq k}^{K} p(k|t)p(l|t) = \sum_{k=1}^{K} p(k|t)(1 - p(k|t)) = 1 - \sum_{k=1}^{K} p^2(k|t) \tag{3.8}$$

$$i_E(t) = -\sum_{k=1}^{K} p(k|t) \log p(k|t) \tag{3.9}$$

The Gini reduction of impurity (Δi_G) and the entropy reduction of impurity(Δi_E) are gained by using i_G, respectively i_E as impurity measures in Equation (3.6).

Prediction   To get a prediction of $\hat{y}$ from a new sample $\mathbf{x}_{new}$ the sample is pushed down each tree of the forest ending in a terminal node. This node is assigned as outcome of the tree to the sample. For regression an average of all tree outcomes is calculated as prediction: $\hat{y}(\mathbf{x}_{new}) = \frac{1}{ntree} \sum_{i=1}^{n} T_i(\mathbf{x}_{new})$. For classification a *majority rule* is suggested:

$$\hat{C} = majority\ rule\{C_i(\mathbf{x}_{new})\} \tag{3.10}$$

where C_i is the class prediction outcome of the ith tree. The majority rule, also denoted as *winner-takes-all* principle, selects the class with the most votes. In the case of RF each tree "votes" for one class.

3.2   Variable Importance

An interesting aspect of RF is that it provides a measurement of the contribution amount of a feature to the classification and prediction accuracy, denoted as *variable importance*. Two methods of calculating the importance are proposed: the *Gini importance*, based on the Gini impurity; and the *permutation accuracy importance*, based on the OOB accuracy [6].

The Gini importance measures the average amount of information gain using the Gini index splitting criterion (Equation 3.8). It is calculated for a given feature variable by averaging the reduction of the Gini impurity (Δi_G) over all trees for those nodes, where the feature variable has been chosen as splitting variable. Since $\Delta i_G(t)$ is calculated as splitting criterion for each node, it just needs to be accumulated during the creation of the RF. The Gini importance is also often denoted as *mean decrease in Gini index*, or *mean decrease in Gini impurity* [6].

Another type of variable importance is calculated from the OOB samples. It measures the prediction strength of each variable. First, for a tree the OOB samples are used for prediction and the accuracy is recorded. Then the samples are perturbed by permuting the variable j. The

accuracy is calculated again for these perturbed samples. By calculating the mean difference of the two accuracies over all trees, the *permutation accuracy importance* for variable j is obtained.

The Gini importance measure correlates well with the permutation accuracy importance [6], although the computational effort for calculating the Gini importance is lower than for the permutation accuracy importance, since it is a side product of the tree creation. On the other hand, Gini importance tends to be biased when predictor variables vary in their number of categories or scale of measurement, whereas the permutation accuracy importance is reliable in such cases, when using sub-sampling without replacement instead of bootstrapping [141].

3.3 Discussion

The Random Forests algorithm, used for regression and classification, has been described in this chapter. The RF algorithm is a supervised learning algorithm, where a prediction model is trained on a given input set containing a p-dimensional feature-vector per input sample, and a output value (target). The basic idea of the RF training algorithm is the growth of many unpruned decision trees with a random selection of input samples per decision tree (bootstrapping) and a random selection of feature variables chosen from the feature-vector for each decision node. In the classification case for each decision node the best feature variable with the optimal splitting threshold is determined by finding the biggest reduction in the node-impurity of the child nodes in an optimization process. The node-impurity measures the amount of "wrong" classes, or entropy in the child-nodes. Two metrics for measuring the impurity have been described in this thesis, the Gini index and the information entropy. It has been shown that the Gini index can be used to determine the importance of a feature variable by calculating the mean decrease in Gini index considering each node using the specific variable as splitting variable. This importance measurement has been denoted as Gini index. A prediction of the output variable for a new input feature vector is done by pushing down the feature variables along each tree until a terminal node is reached. The outcome of the tree is a vote for a class. The final predicted class is obtained by selecting the class with the most votes (majority rule).

CHAPTER 4

Related Work

The previous chapters gave a general overview and background information about the registration and classification methods used in this thesis. The following sections are a more specific overview of related research and state-of-the-art according to the three cornerstones of the thesis: the multimodal breast image registration, the automatic breast lesion segmentation, as well as the automatic lesion classification.

4.1   Multimodal Breast Image Registration

Several methods are proposed for registration of breast DCE-MRI images. A summary is given in the paper of Guo et al. [66]. Rueckert et al. [132] modeled the transformation using a free-form deformation (FFD) based on B-Splines. Normalized mutual information was used as a voxel-based similarity measure. It has been observed by Tanner et al. [146] and Rohlfing et al. [129] that registering DCE-MRI images using free-form deformations tend to shrink the tumor. They developed volume preserving registration methods. Another registration method using finite elements is proposed by Miga [102]. Hill et al. [73] evaluated four methods for registration of DCE-MRI breast images and their influence on the DCE-MRI enhancement curves. They revealed that even a small motion of as little as 1 mm can significantly change the shape of the enhancement curve. Therefore, they suggest that a non-rigid registration method always should be applied before analyzing the contrast enhancement.

Registration of PET images to other modalities is a challenging task, since PET suffers from a low SNR, a low spatial resolution and almost no morphological information [153]. For automatic registration of rigid body parts, like the head, MI-based similarity metrics have been proposed by Collignon et al. [42], Wells III et al. [165] and Maes et al. [96]. A recent evaluation of PET-MRI registration methods for head images is given in [137].

A direct fusion of PET and MRI breast images from separate scanners is only reported with assistance of using fiducial markers [108,153]. An automatic non-rigid image registration of soft tissue PET and MR images without markers has been proposed by Somer et al. [140]. In their

paper they describe a hybrid MRI to PET/CT registration method. They register the MRI image in a first step to the CT image. In a subsequent step the obtained transformation is applied on the PET image to register MRI to PET. They showed that this method is accurate and reliable, and that it gives significant better results than a direct PET to MRI registration. Meanwhile prototypes of combined MRI/PET scanners are available [78], which enables the direct fusion of MRI with PET images [39].

Reiner et al. co-registered full-body DWI and CT scans [126] in their study. They used a software prototype (Multimodality; GE Healthcare) for fusion. Huang et al. [74] used a LDDMM to correct the geometric distortion caused by B0 inhomogeneities in DWI by doing a registration of the b0 DWI to T_2-weighted MRI. To the knowledge of the author no automatic fusion of DWI and MRI breast images has been published.

4.2   Breast Lesion Segmentation

The segmentation of breast lesions can be divided into manual, semi-automatic, and automatic approaches. In a *manual approach* a radiologist delineates the lesion boundary. In this work this annotation represents the gold standard. However, manual segmentation suffers from inter-observer annotation variability, due to interpretation differences of the images [77, 81], and they are time-consuming.

Semi-automatic methods are automatic segmentation approaches that rely on an initial manual setup, such as setting seed points [67], or drawing a ROI [3]. A semi-automatic method, where each pixel/voxel is classified for being cancerous or non-cancerous, has been proposed by Chen et al. [35]. They used a Fuzzy C-Means (FCM) clustering-based method for the segmentation, applied on a breast DCE-MRI series. Wu et al. [170] proposed a Markov Random Field (MRF) based clustering of the time-series data of breast DCE-MRI. An alternative to pixel-wise classification are contour methods, like Active Contour (AC) [79], resulting in a closed contour of the lesion. In Shi et al. [139] an AC is evolved based on FCM clustered DCE-MRI data. A graph-cut based segmentation algorithm has been proposed by Zheng et al. [181]. Recently, Agner et al. [3] presented a hybrid AC method, where the breast DCE-MRI time-series is characterized by principal Eigenvectors derived from Principal Component Analysis (PCA), and the AC evolvement is guided by these eigenvectors.

Automatic segmentation methods do not rely on an initial manual setup, in contrast to semi-automatic methods. Automatic segmentation methods may also be seen as CADe systems, since they automatically localize suspicious regions. Twellmann et al. [152] developed a pixel-wise classification model by using a Support Vector Machine (SVM) classifier in conjunction with DCE-MRI. Woods et al. [169] used 4-D co-occurrence textural features obtained from the breast DCE-MRI to classify the voxels. Yao et al. [174] used co-occurrence and run-length matrix textural features, and applied a wavelet-transformation on the textural temporal breast DCE-MRI sequences to extract frequency features. As classifier they used a SVM committee. Vignati et al. [157] described a method of localizing lesions in fat-suppressed DCE-MRI images. Those images are known to suffer from a low SNR due to the fat-suppression.

Multimodal approaches, combining several modalities, are only reported for CT/PET breast images, where Han et al. [67] segmented lesions by applying a graph-based MRF method on a

	No of lesions (benign/malignant)	AuC	Sensitivity	Specificity	Features
Agner et al. [2], 2011	41 (17 / 24)	0.92	95.00%	82.00%	Morphology + textural kinetics
Chen et al. [36], 2006	121 (44 / 77)	0.85	-	-	Contrast enhancement
Gibbs et al. [60], 2003	40 (17 / 23)	0.92	96.00%	71.00%	Texture, lesion size, time to maximum enhancement, and patient age
Levman et al. [91], 2008	94 (70 / 24)	0.74	62.50%	78.60%	Contrast enhancement
McLaren et al. [100], 2009	71 (28 / 43)	0.82	-	-	Morphology + texture
Meinel et al. [101], 2007	80 (37 / 43)	0.97	-	-	Morphology
Nie et al. [111], 2008	71 (28 / 43)	0.86	-	-	Morphology + texture
Szabó et al. [145], 2004	103 (30 / 73)	0.85	-	-	Morphology + Contrast enhancement
Zheng et al. [181], 2009	36 (14 / 22)	0.97	95.00%	100.00%	Textural kinetics

Table 4.1: Summary of performance of breast DCE-MRI CADx methods.

combined CT/PET image, taking the advantage of both modalities, the high spatial resolution of CT and the functional information of PET. To the author's knowledge no multimodal approach using a combination of DWI, MRI and PET has been proposed.

4.3 Breast Lesion Classification

The aim of breast lesion CADx systems is the automatic classification of a lesion as benign or malignant using a set of features extracted from the lesions; and a classifier that explores the features to discriminate between these two types.

Several methods exploring the DCE-MRI time-series have been proposed [2, 36, 60, 91, 100, 101, 111, 145, 181]. A summary of model-based and model-free parametric DCE-MRI analysis methods is given by Eyal and Degani [51]. Features extracted from the lesion are based on the morphology [101], lesion texture [60, 181], contrast enhancement [36, 91], a combination of morphology and contrast enhancement [145], or a combination of morphology and texture [2, 100, 111]. Table 4.1 summarizes the performance of these methods.

Contrast enhancement methods, like Levman et al. [91], or Chen et al. [36], compute features from the DCE-MRI time-signal curves, such as enhancement ratio and time to peak enhancement. Chen et al. [36] used a FCM clustering method to get characteristic time-signal curves from the lesion. With this method they addressed the problem that drawing manual ROIs within the lesion suffers from significant inter- and intraobserver variability [36]. Szabó et al. [145] combined contrast enhancement features with morphological categories such as margins (smooth, lobulated, irregular, spiculated), or homogeneity (homogeneous, intermediate, heterogeneous). The categories were determined by radiologists. Meinel et al. [101] calculated morphological features from the segmented lesion, like mean volume, area, radial length, spiculation, perimeter length, and compactness. Such morphological features are also used in Agner et al. [2], McLaren et al. [100], and Nie et al. [111], though they combined the morphological features with textural features. Gibbs and Turnbull [60] used the co-occurrence matrix and Haralick features [68] for classifying the lesion. They were able to improve the result by

adding additional morphological (lesion size), and contrast enhancement (time to maximum enhancement) features, as well as the patient age. Chen et al. [37] extended Gibbs and Turnbull's work by replacing the 2-dimensional co-occurrence matrix with a 3-dimensional non-directional co-occurrence matrix. They were able to show that the classification performance of volumetric texture features is significantly better than the classification based on 2-dimensional texture analysis. Zheng et al. [181] and Agner et al. [2] proposed spatiotemporal texture features, denoted by Agner et al. as *textural kinetics*. These features aim in capturing the spatiotemporal changes in breast lesion texture. Zheng et al. [181]computed the discrete Fourier transform of Gabor filtered texture features of the DCE-MRI series to obtain the spatiotemporal changes. Agner et al. [2] combined spatiotemporal texture features such as Gabor filter, Sobel filter, first-order statistics, and second-order (Haralick) statistics, with morphological and contrast enhancement features.

Using DWI for classification shows promising results [24, 65, 97, 171–173]. Marini et al. [97] report a sensitivity of 80% and a specificity of 81% using only DWI for classification. Yabuuchi et al. [171, 172] combined DWI with DCE-MRI morphological categories and contrast enhancement features. They report a sensitivity, specificity, and accuracy of 92%, 86%, and 91% for mass-enhanced lesions [172], and 87%, 86%, and 86%, respectively for non-mass-like enhanced lesions [171].

To the authors knowledge using combined DCE-MRI, PET, and DWI modalities in a CADe or CADx system is a novel approach.

4.4 Discussion

In this chapter a state-of-the-art summary has been given for the three cornerstones of the thesis: the multimodal breast image registration, the breast lesion segmentation and classification. References to methods for DCE-MRI registration, PET to MRI registration via CT, and DWI to MRI registration have been given. Semi-Automatic and automatic segmentation methods have been listed, either aiming on a segmentation on a voxel level, by classifying each voxel of being lesion or background, or aiming on considering the topology by using active contours or graph cut methods. In addition, methods for the classification of the lesion have been presented in this chapter. The performance of methods analyzing the DCE-MRI time-series have been listed and the approaches used for getting discriminative features, like morphology, contrast enhancement, and textural changes, have been described. Finally, promising results of using DWI for classification has been summarized in this chapter.

Methodology

In this chapter the CAD pipeline is presented in detail, from data acquisition, registration and segmentation to classification. In the first section an overview of the CAD pipeline is given. The second section presents the medical imaging methods used for the CAD system. Furthermore, the image acquisition and preprocessing steps are elucidated in this section. The third section covers the multimodal image registration method. In Section 5.4 the extraction of the features needed for the segmentation process is described. Section 5.5 explains the segmentation process. Section 5.6 involves the extraction of features needed for the classification of the lesion. In the final section the classification method is described.

Notations   For the subsequent formulas following notation is used: The three-dimensional images are denoted as $I_{modality}$, where I_{dce} is the DCE-MRI modality, I_{dwi} is the DWI modality, and I_{pet} is the PET modality. The image functions are parameterized by the spatial voxel coordinates $\mathbf{x} = (x\,y\,z)$, with $x, y, z \in \mathbb{N}^+, x \le X, y \le Y, z \le Z$. The constants X, Y, Z represent the size of the image in each dimension.

$I_{modality}(\mathbf{x})$ returns the intensity value of the modality at the given coordinates. The DCE-MRI modality is indexed by an additional parameter $i \in \mathbb{N}^+$, with $1 \le i \le N$ and N is the number of frames in the DCE-MRI sequence. So $I_{dce}(\mathbf{x}, i)$ specifies the intensity value at spatial position $\mathbf{x}$ for the DCE-MRI frame i.

5.1   The Computer Aided Detection & Diagnosis (CAD) Pipeline

The CAD pipeline is implemented as a process of six steps (Figure 5.1). The first step is the data acquisition and preprocessing step. The multimodal images are obtained from a MRI and a PET/CT scanner, converted to a compact image file format, truncated and finally converted to a unified image coordinate system. As a second step the images are co-registered and scaled to a reference image, so that they are all aligned and all of them have the same spatial resolution. After the registration process the feature extraction takes place. Features needed for the segmentation process are calculated for each voxel within the breast. The segmentation is implemented
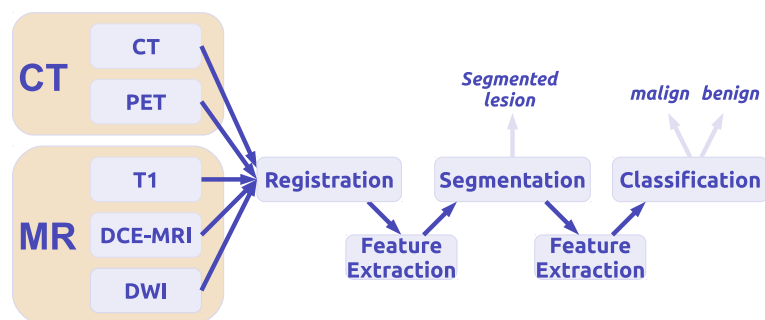
Figure 5.1: A representation of the CAD pipeline.

as a binary classification process, where a RF machine learning algorithm decides, depending on the voxel features, if a voxel is classified as cancerous tissue or normal tissue. As a result of this step a segmentation of the lesion is achieved. From the segmented lesion features are extracted, which are necessary for the classification of the lesion. These features are calculated from regions or clusters of voxels, in contrast to the segmentation features, which are voxel-based. The classification in benign and malign lesions is done by a RF machine learning algorithm.

5.2   Image Acquisition

In the image acquisition step three dimensional images are acquired from a 3 Tesla MRI scanner, and a combined PET/CT scanner. At the MRI scanner the DCE-MRI and the DWI sequences are recorded in a single session. All MR images are acquired from a 3 Tesla MRI scanner (Tim Trio, Siemens, Erlangen, Germany) using a 4-channel breast coil in vivo. At the PET/CT scanner an image of the thorax is acquired for each modality, where the PET and CT images are initially aligned by the scanner software.

5.2.1   Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI)

In the DCE-MRI acquisition sequence a combination of high spatial resolution and high temporal resolution MR images are recorded, where the high spatial resolution images are used to capture the detailed morphological information of the lesion, and the high temporal resolution images capture the contrast enhancement over time. A summary of the DCE-MRI acquisition protocol is given in the following part of this section. The fully detailed acquisition protocol, as developed by Pinker et al., is described in [119].

The DCE-MRI acquisition protocol consists of 5 parts. First, a high spatial resolution T1-weighted image with water excitation and fat suppression is recorded with an acquisition time of 2 minutes. The dimension of the image is 320x320x120, with a voxel size of 1 mm isotropic. This image is hereinafter referred to as *DCE-MRI Pre-Contrast image*, and in the mathematical context as $I_{dce\text{-}pre}$. Subsequently, 16 contrast-enhanced T1-weighted images with an acquisition time of 13 seconds are recorded. The higher temporal resolution used in this sequence leads to

Figure 5.2: DCE-MRI recording sequence in chronological order. The time in this diagram is the time needed for recording the image or image sequence.

a lower spatial resolution of 1.7 mm isotropic and an image dimension of 192x192x72. After 75 seconds a contrast agent is injected. So the first 5 images of the sequence are recorded without contrast agent, and are used to determine the baseline of the MR image. Next, a high spatial resolution T1-weighted image is acquired. At this time the peak of the contrast enhancement at the lesion is expected [86]. This high-resolution image is referred to as *DCE-MRI Peak-Contrast image*, respectively $I_{dce\text{-}peak}$. Thereafter, 24 images with a high-temporal resolution are acquired, to capture the washout of the contrast agent. Finally, a high spatial resolution image is acquired, which is referred to as *DCE-MRI Post-Contrast image*, respectively $I_{dce\text{-}post}$. The DCE-MRI acquiring sequence is illustrated in Figure 5.2. Applying this protocol results in a set of three MR images with a high spatial resolution, two of them contrast enhanced, and an image sequence of 40 MR images, denoted as I_{dce}, with a high temporal resolution, capturing the contrast uptake and washout.

5.2.2 Diffusion Weighted magnetic resonance Imaging (DWI)

The DW images are acquired in the same recording session as the DCE-MR images. A combined b value protocol of 0 and 850 sec/mm2 is used, resulting into two DW images and an ADC mapping calculated from the two DW images. In Figure 1.4 the DWI and ADC images acquired from a patient are shown. A novel DWI sequence is used in the acquisition process, which is less distorted and which has a higher SNR than other DWI sequences [123]. Details of the DWI protocol and its parameters are listed in [25].

The resolution of the image is 2.09×2.09 mm, with a slice thickness of 5.5 mm, resulting in an image dimension of $172 \times 86 \times 24$. The first DW image with a b value of 0 sec/mm2 is used in the subsequent registration and segmentation steps. This image has a higher contrast and there are more morphological structures visible than in the second DW image using the b value of 850 sec/mm2, as seen in the Figure 1.4 presented in the Introduction chapter. In subsequent parts the DW image with a b value of 0 sec/mm2 is referred as DWI (I_{dwi}), and the DWI-ADC mapping is denoted as ADC, respectively I_{adc}.

5.2.3 Positron Emission Tomography (PET)/Computed Tomography (CT)

PET and CT images of the thorax are acquired in a combined PET/CT scanner. $[^{18}F]$ *fluorodes-oxyglucose ($[^{18}F]$ FDG)* is used as a radiotracer for the PET image. The spatial resolution of the PET image is 4×4 mm with a slice thickness of 3 mm. The dimension of the image is $168 \times 168 \times 74$. The CT image has a resolution of 1.37×1.37 mm, a slice thickness of 3 mm

and an image dimension of $512 \times 512 \times 74$. The PET and CT images are aligned automatically by the scanner software.

5.2.4 Image Preparation

Following pre-processing steps are performed after image acquisition:

Conversion The acquired images are initially stored in a Picture Archiving and Communication System (PACS), from which they have been exported for further processing using the DICOM [23] image format. The DICOM format stores each slice of each acquired image in a separate file, resulting in directories containing several thousand DICOM files for each patient. The DICOM images are therefore converted to a more compact medical image file format, the Neuroimaging Informatics Technology Initiative (NIfTI) (`http://nifti.nimh.nih.gov/`) image format. In this format one file is representing one volumetric dataset. Hence, one single NIfTI file per image modality is gained after conversion.

Mask creation A binary mask $M = \{\mathbf{x} | \mathbf{x} \in breast\}$ is created from $I_{dce\text{-}pre}$, to segment the region inside the breast from the surrounding air. All subsequent calculations of features and performance ratings are done for voxels specified inside the breast only. A single mask for all modalities is calculated, since all modalities are registered in the subsequent step to $I_{dce\text{-}pre}$. M is created by using an intensity based region-growing algorithm [1]. The seed points are defined manually by selection points within the surrounding air. The intensity threshold is also defined manually, so that the region growing algorithm selects air voxels only, and stops at the soft-tissue border of the breast. The mask is inverted afterwards, resulting in a selection of the breast region. Small errors in the segmentation, which occurred due to the partially low contrast of the skin-tissue, have been manually corrected.

Truncating PET/CT The original CT and PET images show the whole thorax, whereas the MR images cover the breast region only. For this reason the PET and CT images are truncated by manually selecting a bounding box in a way, that the field of view is approximately the same than the field of view of the MR images.

   The NIfTI file format has a coordinate system included, to represent the position, orientation and resolution of the underlying image. It is used i.e. for converting image coordinates to patient coordinates and vice versa. The subsequent registration process uses this information too. To get initially roughly aligned CT/PET and MR images, the coordinate system origins of the CT and PET images are replaced by the one from the MR image. Figure 5.3 illustrates the truncating process and the initial alignment of the CT and MR image. The truncated PET image is subsequently denoted as I_{pet}, and the CT image as I_{ct}.

   In Table 5.1 a summary of all acquired images and their denotions is given.

40
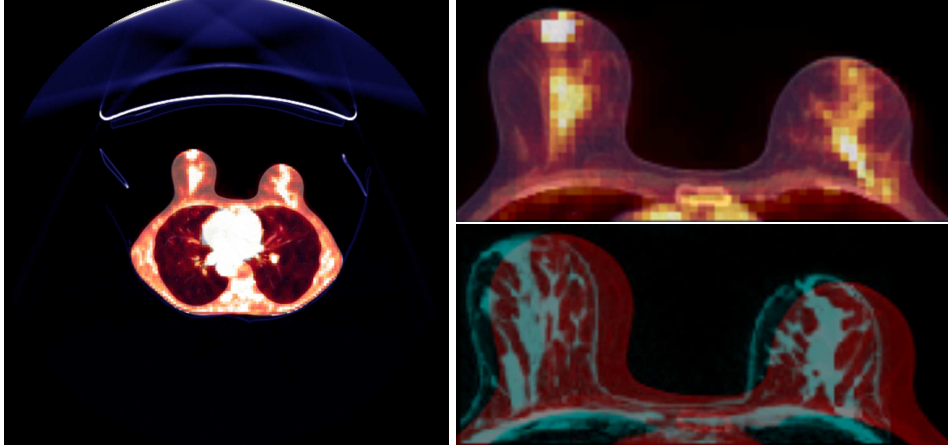
Figure 5.3: Left: PET and CT image of the thorax as acquired from the PET/CT scanner. Top right: Truncated PET and CT image. Bottom right: Overlay of the MRI (cyan) and the truncated CT image (red) using the MRI coordinate system origin.

Image type	Denoted as	Description
DCE-MRI	$I_{dce}(\mathbf{x}, i)$	DCE-MRI time-series (40 images)
DCE-MRI Pre-Contrast	$I_{dce\text{-}pre}(\mathbf{x})$	DCE-MRI high-resolution pre-contrast image
DCE-MRI Peak-Contrast	$I_{dce\text{-}peak}(\mathbf{x})$	DCE-MRI high-resolution peak-contrast image
DCE-MRI Post-Contrast	$I_{dce\text{-}post}(\mathbf{x})$	DCE-MRI high-resolution post-contrast image
DWI	$I_{dwi}(\mathbf{x})$	DWI (b-value: 0 sec/mm2)
ADC	$I_{adc}(\mathbf{x})$	DWI ADC mapping
PET	$I_{pet}(\mathbf{x})$	Truncated PET image
CT	$I_{ct}(\mathbf{x})$	Truncated CT image
Mask	$M(\mathbf{x})$	Binary mask image obtained from $I_{dce\text{-}pre}$

Table 5.1: Summary of image types obtained in the data acquisition and pre-processing step.

5.3 Registration of Image Modalities

The registration of the image modalities is the next step in the CAD pipeline, with the objective of transforming all images to a reference coordinate system. The image modalities are aligned using a 3D voxel intensity-based approach.

The registration is done in two steps. First an affine registration is performed, to do a global alignment of the breast images, followed by a non-rigid transformation. The latter transformation is necessary to compensate the patient movement and breathing during a DCE-MRI recording sequence, the distortions, which are inherent in DWI, as well as the different image acquiring conditions and patient positions in the MRI and PET/CT scanner. For instance, all MRI modalities are recorded by placing a coil between the two breasts, causing a slight deformation, which is not present at the PET/CT acquisition, and which needs to be compensated in the registration process. $I_{dce\text{-}pre}$ is used as the reference image, to which all other images are
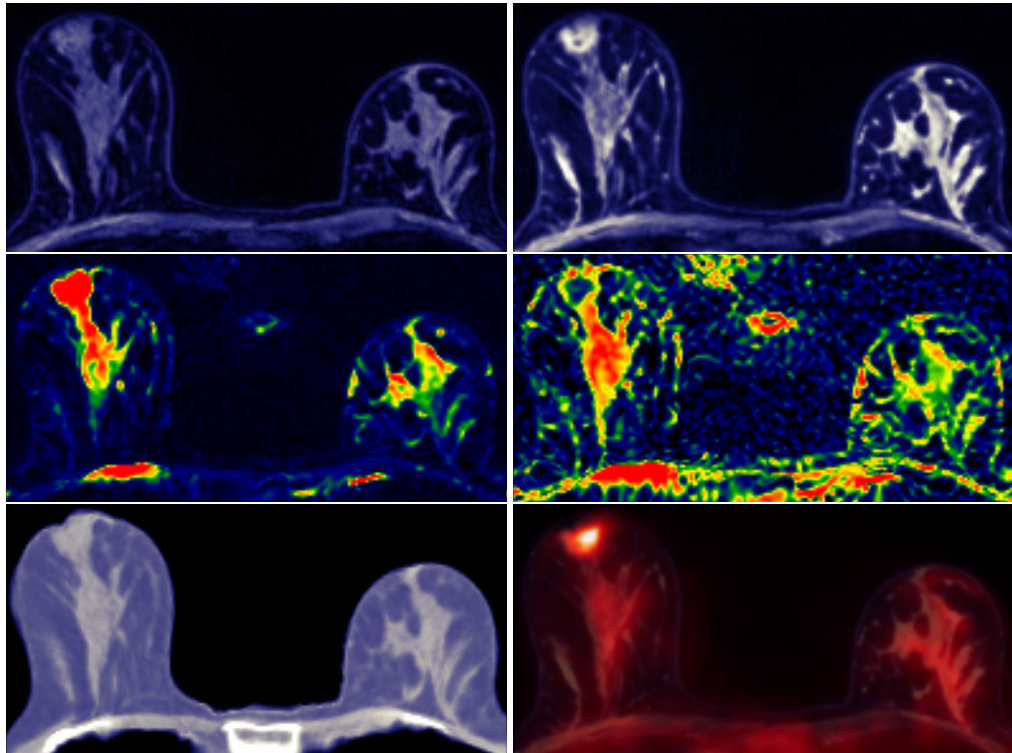
Figure 5.4: Results of the registration process for one patient. Top row: Reference Pre-Contrast MRI and registered Post-Contrast MRI; middle row: registered DWI and ADC map; bottom row: registered CT and PET images. The PET image is overlayed on the MRI image to visualize the combined morphologic and functional information of these two modalities.

registered to. All images are upscaled to the $I_{dce\text{-}pre}$ resolution after registration using a trilinear interpolation. After the registration step all image modalities share a common coordinate system, they are aligned, and they have the same orientation and resolution. After registration the spatial coordinate vector $\mathbf{x}$ refers to the same spatial location in all image modalities. Figure 5.4 shows the result of the registration process for one patient.

Registration Framework   The ANTS framework by Avants et al. [14] has been chosen for registering the multimodal images. It implements several transformation models, like SyN, Diffeomorphic Demons, elastic and rigid/affine registrations, as well as several kinds of similarity metrics, like MI, CC, and MSQ. ANTS is open source, it is well documented and it works in Linux, Windows and Mac OsX environments. These facts as well as the top rankings of SyN in the Klein et al. [82] study and the EMPIRE10 challenge [109] is the main motivation for using this framework.

Affine Registration   At this registration step an affine transformation matrix $\mathbf{T}^{Aff}$ is successively modified in an optimization process. The registration is performed on a multi-resolution

scale based on a 3 level Gaussian pyramid, as Avants et al. [14] suggests for images with 1 mm^3 resolution. The initial registration begins with a quarter of the original resolution, the second turn uses half of the resolution and the final turn is performed with the full resolution. MI with 32 bins is used as the similarity metric.

Nonrigid Registration During the non-rigid registration step a deformation map $\mathbf{T}^{SyN}$ is created and successively optimized, covering the non-rigid deformation of the breast. As similarity metric CC is used for registering I_{ct} and I_{dwi} to $I_{dce\text{-}pre}$. The I_{dce} series is registered using the MI metric. As transformation model GrSyN is used. Other metrics and transformation models have been evaluated, whereas these combinations of metrics and transformation models have the best performance. Details of the evaluation process are found in Section 6.1 of the "Experiments" chapter.

The non-rigid optimization process is also performed using a 3 level Gaussian pyramid. The maximum number of iterations is limited to 500, 300 and 150 in respect to the resolution level.

The registration of DWI to MRI can be formally written as:

$$I_{dwi}^{*}(\mathbf{x}) = I_{dwi}(\mathbf{T}_{dwi,mri}(\mathbf{x})) \tag{5.1}$$

where $\mathbf{T}_{dwi,mri}$ is the transformation obtained in the registration process when registering I_{dwi} to $I_{dce\text{-}pre}$. This transformation is a composition of the nonrigid and the affine transformation: $\mathbf{T}_{dwi,mri}^{SyN}(\mathbf{T}_{dwi,mri}^{Aff}(\mathbf{x}))$. The registration transformations of the other modalities I_{ct}^{*}, I_{dce}^{*}, $I_{dce\text{-}peak}^{*}$, and $I_{dce\text{-}post}^{*}$ are defined analogous. Since I_{adc} and I_{dwi} share the same coordinate system, $\mathbf{T}_{dwi,mri}$ is used for the transformation of I_{adc} to the reference image.

PET to MRI registration I_{pet} is not directly registered to $I_{dce\text{-}pre}$. The low resolution, together with the absence of morphological information in the PET modality, severely complicates an accurate non-rigid registration. Therefore the PET image is registered in an indirect way to the MR image by registering the CT image to the MR image in a first step, and applying the registration transformation $\mathbf{T}_{ct,mri}$ on the PET image in a second step:

$$I_{pet}^{*} = I_{pet}(\mathbf{T}_{ct,mri}(\mathbf{x})) \tag{5.2}$$

The higher resolution of the CT image, and the presence of morphological structures within the CT image results in an accurate registration of I_{ct} to $I_{dce\text{-}pre}$, and as a consequence the accurate registration of I_{pet} to $I_{dce\text{-}pre}$. This method is inspired by Somer et al. [140], where soft-tissue PET/CT and MR images have been registered in a similar way.

After registration the images I_{dwi}^{*}, I_{adc}^{*}, I_{pet}^{*}, I_{ct}^{*}, I_{dce}^{*}, $I_{dce\text{-}peak}^{*}$, $I_{dce\text{-}post}^{*}$, and $I_{dce\text{-}pre}$ share a common reference coordinate system. Subsequently the spatial coordinates $\mathbf{x}$ refers to this coordinate system. To facilitate the reading, the * is skipped in the subsequent sections and chapters.

5.4 Data Preparation and Segmentation Feature Extraction

After image registration the data is pre-processed and features are extracted for the subsequent segmentation step. The preprocessing step consists of the annotation of the lesion and the normalization of the MRI data.

Annotation The annotation of the lesion is used as a reference for training and as gold-standard for validation of the segmentation process. The annotation is a labeling $L(\mathbf{x}) \in \{0, 1\}$ operating in the reference frame, where the non-lesion regions are labeled as 0, and the lesion voxels are labeled as 1. An experienced radiologist manually annotated the lesion either on the registered $I_{dce\text{-}peak}$ or $I_{dce\text{-}post}$, depending on where the lesion is better visible. ITK-Snap [178] is used as the annotation tool. The lesions are annotated manually by paintbrushing and by using active snake segmentation [79], which is driven by image intensity. Care has been taken, that the whole lesion is covered by the annotation. Although, especially at the border of the lesion partial volume effects are taking place, rendering the decision of a voxel imprecise for being cancerous or non-cancerous. Hence, inaccuracies in the annotation at the lesion border are possible.

Data normalization The variable scale of the signal intensities is inherent in MR images and varies from patient to patient. Therefore the data is normalized to get a uniform scale for all patients MR images. As normalization the standard score is used: $I'(\mathbf{x}) = \frac{I(\mathbf{x}) - \hat{\mu}}{\hat{\sigma}}$. $\hat{\mu}$ and $\hat{\sigma}$ are estimated for each patient from the masked Pre-Contrast MR image intensities $I_{dce\text{-}pre} \cap M$. These precontrast parameters are also used for normalizing I_{dce}, $I_{dce\text{-}peak}$, and $I_{dce\text{-}post}$. The contrast uptake, which differs for each patient, would have an influence on the statistic parameters $\hat{\mu}$ and $\hat{\sigma}$ and the normalization, if they would have been calculated separately for each frame of the DCE-MRI series. Therefore the statistical parameters of $I_{dce\text{-}pre}$ are used for normalization of the contrast enhanced images, assuming that all images of the DCE-MRI sequence share the same scale as $I_{dce\text{-}pre}$. Since for segmentation and classification only the normalized values are used, in the subsequent sections I_{dce}, $I_{dce\text{-}pre}$, $I_{dce\text{-}peak}$, and $I_{dce\text{-}post}$ refers to the normalized images.

5.4.1 Segmentation Feature Extraction

Subsequently the features needed for the segmentation process are extracted from the preprocessed data on a voxel base. The features can be categorized into following groups:

- Signal intensity based features,

- texture based features, and

- DCE-MRI kinetic curve features.

In the first category features are subsumed, which are based on the signal intensities of the image modalities; the features from the second category are used to analyze textural properties of the lesion; and the features from the third category are describing properties of the kinetic curve of the DCE-MRI sequence.

A list of all features extracted for segmenting the lesions is given in Table 5.4 at the end of this section. A detailed performance evaluation of the segmentation feature is given in Section 6.2 of the "Experiments" chapter. Subsequently, the extracted features at a given spatial position $\mathbf{x}$ are represented by a feature vector $\mathbf{f}_{type}(\mathbf{x})$, where $type$ specifies the category, respectively modality of the extracted feature.

Signal intensity based features The signal intensities from I_{dwi}, I_{adc}, I_{pet}, $I_{dce\text{-}pre}$, $I_{dce\text{-}peak}$, and $I_{dce\text{-}post}$ are extracted from each voxel within the mask M. Formally, $f_{dwi}(\mathbf{x}) = I_{dwi}(\mathbf{x})$, analogous for the other modalities. Also the signal intensities of I_{dce} are extracted, resulting in 40 values per voxel, representing the signal intensity change over time: $f_{dce}(\mathbf{x}) = (I_{dce}(\mathbf{x}, 1), ..., I_{dce}(\mathbf{x}, N))$. In Figure 5.5 the DCE-MRI signal intensity values for one voxel within the lesion, one voxel from a fat tissue, and one voxel from a lobules region is plotted over time. An intensity change at the cancerous tissue can be observed beginning with frame 7. At this time point the contrast agent causes a change in the signal intensity. Intensity values after this time point are discriminative features to distinct between cancerous and non-cancerous voxels, whereas the intensity values of the pre-contrast time-points are not able to discriminate between cancerous and lobules tissue.
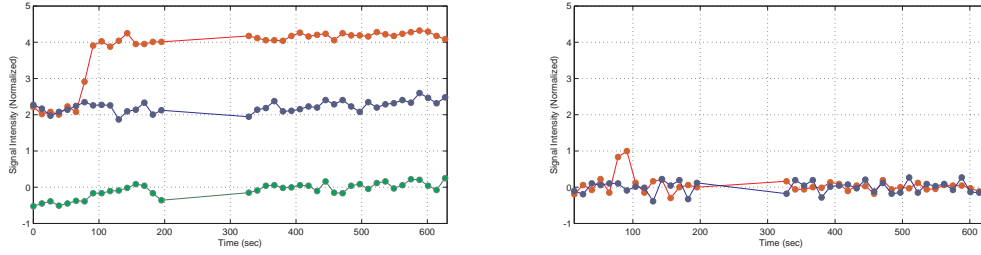
Figure 5.5: Left: Signal intensity of a voxel from a DCE-MRI time series. Right: Derivation of the signal intensity values. Red: voxel from a cancerous tissue region, blue: voxel from a lobules region, green: voxel from a fatty tissue region.

The rapid contrast uptake in cancerous tissue after injection of the contrast agent is modeled by using an approximate derivative over time of I_{dce}, denoted as ΔI_{dce}. The derivative at spatial position $\mathbf{x}$ is calculated by using the forward difference method:

$$\Delta I_{dce}(\mathbf{x}, i) = \frac{I_{dce}(\mathbf{x}, i+1) - I_{dce}(\mathbf{x}, i)}{t_{i+1} - t_i}, \forall 1 \leq i < N - 1 \tag{5.3}$$

where t_i is the acquisition time point in seconds of frame i. The normalization term $t_{i+1} - t_i$ is not necessary, if the acquiring time points are equally spaced. This is not the case for the DCE-MRI protocol used in this CAD pipeline, where a 2 minute gap exists between frame 16 and 17. At this time-point $I_{dce\text{-}peak}$ is acquired. Therefore the normalization term has been added to the calculation. The derivative intensity values are plotted in Figure 5.5. A peak in the derivative intensity values for cancerous tissue can be observed at the time where the contrast agent is injected and the signal intensity changes rapidly. This peak is not present for non-cancerous

tissues, which are not affected by the contrast enhancement. The feature vector $f_{\Delta dce}$ contains the intensities of ΔI_{dce}.

The difference in signal intensity of $I_{dce\text{-}peak}$ to $I_{dce\text{-}pre}$ and $I_{dce\text{-}post}$ to $I_{dce\text{-}pre}$ are used as additional features. Formally written: $f_{\Delta peak\text{-}pre}(\mathbf{x}) = I_{dce\text{-}peak}(\mathbf{x}) - I_{dce\text{-}pre}(\mathbf{x})$, and $f_{\Delta post\text{-}pre}(\mathbf{x}) = I_{dce\text{-}post}(\mathbf{x}) - I_{dce\text{-}pre}(\mathbf{x})$.

Texture based features A 3-dimensional Gray-Level Co-occurrence Matrix (GLCM) and second-order statistics based on this matrix as suggested by Haralick [68] are used to describe textural properties of the DCE-MRI. Using co-occurrence textural features is inspired by the work of Woods et al. [169], where spatiotemporal texture features are used for segmentation of the lesion, as well as the work of Chen et al. [37], where non-directional 3D GLCMs are used to classify the lesions.

The GLCM is a way to describe the distribution of intensities within a region, as well as spatial variations of intensities in a given direction. It is defined as following: Consider an image with L possible grey-levels. Let $\mathbf{d}$ be a displacement vector describing the difference in the spatial locations of two voxels. The entry (i, j) of a co-occurrence matrix $\mathbf{G_d}$ is the number of times that intensities of i and j occurs in voxel-pairs having a spatial distance of $\mathbf{d}$. The size of $\mathbf{G_d}$ is $L \times L$.

To get the empirical probability of occurrence of voxel-pairs with an intensity of i and j, the co-occurrence matrix is normalized by the total number of voxel-pairs:

$$\mathbf{p_d}(i,j) = \frac{\mathbf{G_d}(i,j)}{\sum_{k,l=1}^{L} \mathbf{G_d}(k,l)} \tag{5.4}$$

The displacement vector $\mathbf{d}$ can be interpreted as a directional vector defining the offset between two voxel. For example in the two dimensional case a displacement vector of $(1, 0)$ pairs the pixel with the one immediately to the right, whereas $(0, -2)$ is the pixel-pair of the pixel two pixels above of the current pixel. For a certain distance d, there are 8 neighboring voxel-pairs in 4 independent directions. In the 3-dimensional case there are 26 neighboring voxel-pairs in 13 independent directions. The displacement vectors describing the independent directions are summarized in Table 5.2.

From the GLCM thirteen second-order statistics as proposed by Haralick [68] are computed (f_1 to f_{13}). These statistics are commonly referred to as *Haralick Features*. They describe properties of the texture like contrast, entropy or homogeneity. A detailed description of all 13 Haralick texture features is given in the Appendix (Table A.1).

For every voxel inside the breast a $5 \times 5 \times 5$ window is used for calculating the GLCM matrix. On the one hand this small window size has been chosen to cover small texture regions, like small lesions. On the other hand, if the window size is too small, the statistical expressiveness would have been reduced. Border voxels are skipped in the calculation, assuming that there is no lesion localized in the voxel border. The intensity values are binned to 32 gray-levels, as proposed by Chen et al. [37]. The distance d is set to 1 voxel and a GLCM is calculated for all 13 independent directions. A non-directional mean GLCM is computed from these 13 matrices and it is normalized afterwards. The thirteen Haralick features are calculated from this non-directional GLCM.

46

2-dimensional	3-dimensional
d	**d**
$(d, 0)$	$(d, 0, 0)$
$(0, d)$	$(0, d, 0)$
(d, d)	$(0, 0, d)$
$(d, -d)$	$(d, d, 0)$
	$(-d, d, 0)$
	$(0, d, d)$
	$(0, d, -d)$
	$(d, 0, d)$
	$(d, 0, -d)$
	(d, d, d)
	$(-d, d, d)$
	$(d, d, -d)$
	$(-d, d, -d)$

Table 5.2: Displacement vectors **d** covering all unique directions for 2-dimensional and 3-dimensional images. d is the offset between the voxels

The Haralick features are calculated for I_{dce}, as well as for $I_{dce\text{-}pre}$, $I_{dce\text{-}peak}$, and $I_{dce\text{-}post}$. To reduce the computational effort and the memory consumption, every second DCE-MRI frame is skipped in the calculation, resulting in Haralick features for 20 frames:

$$f_{tex\text{-}dce}(\mathbf{x}) = (\mathsf{f}_{1,1}(\mathbf{x}), \mathsf{f}_{1,2}(\mathbf{x}), ..., \mathsf{f}_{1,13}(\mathbf{x}), \mathsf{f}_{2,1}(\mathbf{x}), ..., \mathsf{f}_{N,13}(\mathbf{x})) \tag{5.5}$$

where in this case $\mathsf{f}_{i,j}$ denotes the Haralick feature j of frame i, obtained from I_{dce}. Analogous, the texture feature of the high-resolution MR images is:

$$
\begin{aligned}
f_{tex\text{-}mri}(\mathbf{x}) = &(\mathsf{f}_{pre,1}(\mathbf{x}), \mathsf{f}_{pre,2}(\mathbf{x}), ..., \mathsf{f}_{pre,13}(\mathbf{x}), \\
&\mathsf{f}_{peak,1}(\mathbf{x}), ..., \mathsf{f}_{peak,13}(\mathbf{x}), \mathsf{f}_{post,1}(\mathbf{x}), ..., \mathsf{f}_{post,13}(\mathbf{x}))
\end{aligned}
\tag{5.6}
$$

$\mathsf{f}_{pre,j}$ denotes the jth Haralick feature obtained from $I_{dce\text{-}pre}$.

The change in the texture over time due to the contrast enhancement in the DCE-MRI time series is modeled by using the derivative of the Haralick features. The derivative is approximated by forward differencing, analogous to ΔI_{dce}:

$$\Delta \mathsf{f}_{i,j} = \frac{\mathsf{f}_{i+1,j} - \mathsf{f}_{i,j}}{t_{i+1} - t_i}, \tag{5.7}$$

$$f_{\Delta tex\text{-}dce} = (\Delta \mathsf{f}_{1,1}, ..., \Delta \mathsf{f}_{N-1,13}) \tag{5.8}$$

DCE-MRI Kinetic Curve Features   This kind of features describes kinetic properties of the DCE-MRI time-signal intensity curve. Following properties are calculated for each voxel:

- Area under Curve (AuC),

- Maximum of intensity Enhancement (MaxEnh),

- Maximum of Slope (MaxSlope) and

- Time to Half Maximum (THM).

These features are calculated from the relative change of the intensity to the baseline intensity in I_{dce}, the enhancement rate. The baseline S_0 for a voxel is determined from the four normalized precontrast DCE-MR images, by calculating the mean intensity of these four intensities:

$$S_0(\mathbf{x}) = \frac{1}{4} \sum_{i=1}^{4} I_{dce}(\mathbf{x}, i) \tag{5.9}$$

The enhancement rate R_t is then calculated by following formula:

$$R_t(\mathbf{x}, i) = (I_{dce}(\mathbf{x}, i) - S_0(\mathbf{x}))/S_0(\mathbf{x}), \ 1 \le i \le N \tag{5.10}$$

From R_t the kinetic curve features are determined. The AuC is calculated by using a trapezoidal integration approximation. MaxEnh is determined by finding the maximum intensity in R_t. MaxSlope describes the steepest slope along the time-intensity curve and is defined as the maximum of the forward difference of R_t. And THM is defined as the time-point, where the intensity value exceeds the half of the MaxEnh. The formal definition of the features is given in Table 5.3. Figure 5.6 illustrates the extracted features from a voxel of DCE-MRI time-intensity curve. These features are subsumized in the feature vector $f_{kin}(\mathbf{x}) = (AuC, MaxEnh, MaxSlope, THM)$.
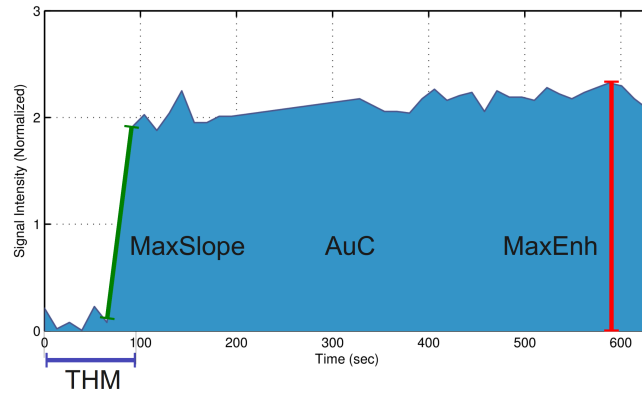
Figure 5.6: Illustration of the four features extracted from a voxel of the DCE-MRI time series. Red: Maximum Enhancement; Green: Maximum Slope; Magenta: Time To Half Maximum; Blue: Area under Curve

Feature	Description	Calculation
AuC	Area under Curve	$AuC(\mathbf{x}) = \sum_{i=1}^{N-1} (t_{i+1} - t_i) \dfrac{R_t(\mathbf{x}, i) + R_t(\mathbf{x}, i+1)}{2}$
MaxEnh	Maximum Enhancement	$MaxEnh(\mathbf{x}) = \max(R_t(\mathbf{x}, i)); \; 1 \leq i \leq N$
MaxSlope	Maximum Slope	$MaxSlope(\mathbf{x}) = \max\left(\dfrac{R_t(\mathbf{x}, i+1) - R_t(\mathbf{x}, i)}{t_{i+1} - t_i}\right)$
THM	Time to Half Maximum	$THM(\mathbf{x}) = \min(t_i); \;\; \text{subject to } R_t(\mathbf{x}, i) \geq MaxEnh/2$

Table 5.3: Formal definition of the DCE-MRI kinetic curve features f_{kin}. $R_t(\mathbf{x}, i)$ is the relative intensity change to the baseline intensity at spatial position $\mathbf{x}$ for frame number i. t_i is the acquisition time-point of image i relative to the time-point of acquisition of the first image in seconds.

5.5 Segmentation of the Lesion

In the segmentation step each voxel within the breast is classified as either cancerous or non-cancerous by a Random Forests (RF), resulting in a segmentation labeling $l(\mathbf{x}) \in \{0, 1\}$, where 1 denotes a lesion voxel and 0 background. The classification is based on a feature vector $f(\mathbf{x})$, which is a subset of the previously extracted features. For instance, a feature vector $f(\mathbf{x})$ containing the modalities I_{pet} and I_{dwi} has the format: $f(\mathbf{x}) = (f_{pet}, f_{dwi})$. The feature vector is used for training and prediction of new cases. Several feature subsets have been evaluated, where the intensity based features f_{pet} and f_{dwi} in combination with f_{dce} or Δf_{dce} showed the best performance. A detailed evaluation of combination of feature subsets is given in the "Experiments" chapter (Section 6.2). Figure 5.7 illustrates the process of the segmentation.

RF Training In order to get a prediction model for a given feature subset, the RF needs to be trained on this subset. For this reason a training feature set f_{train} is evolved using the manual annotated lesions $L^i(\mathbf{x})$ as target features and $f^i(\mathbf{x})$ as training input data, where i is in this case the patient index. To get a reasonable size of training data, *random undersampling* [70] is performed by sampling randomly 1000 voxels per patient i from the annotated lesion, as well as

Feature vector	Description	No. of Features
f_{dwi}	DWI intensity value	1
f_{adc}	Apparent diffusion coefficient value	1
f_{pet}	PET intensity value	1
$f_{dce\text{-}pre}$, $f_{dce\text{-}peak}$, $f_{dce\text{-}post}$	Intensity values of the high-resolution MR images	3
$f_{\Delta peak\text{-}pre}$	Change in intensity between MRI Peak-Contrast and MRI Pre-Contrast	1
$f_{\Delta post\text{-}pre}$	Change in intensity between MRI Pre-Contrast and MRI Post-Contrast	1
f_{dce}	DCE-MRI intensity values for each frame of the DCE-MRI series.	40
$f_{\Delta dce}$	Difference of DCE-MRI intensity values between two subsequent frames	39
$f_{tex\text{-}dce}$	13 Haralick features calculated for every second frame of the DCE-MRI series	13*20
$f_{\Delta tex\text{-}dce}$	Difference of the 13 Haralick features between two subsequent frames	13*19
$f_{tex\text{-}mri}$	13 Haralick features calculated from the high-resolution MR images	13*3
f_{kin}	Kinetic Curve features: Area under Curve, Maximum Enhancement of Intensity, Maximum of Slope, and Time to Half Maximum	4
	Sum:	637

Table 5.4: List of all features extracted from a voxel for the segmentation classification. Features are grouped by 'Signal Intensity Features', 'Texture Based Features' and 'DCE-MRI Kinetic Curve Features'.

5000 voxels from non-lesion tissues. The higher amount of non-lesion tissue samples has been chosen to cover the higher variability of tissues in these areas. Details of the sampling process are given in the Experiments chapter - Section 6.2. The training feature set f_{train} is then used to train a prediction model to determine between lesion tissue and background.

RF Classification   To get the segmentation labeling $l^n(\mathbf{x})$ for a new patient with index n each voxel $\mathbf{x}$ of the breast is classified by the trained RF using the same features as in the training step (e.g. $f^n(\mathbf{x}) = (f^n_{pet}(\mathbf{x}), f^n_{dwi}(\mathbf{x}))$). The classification is based on the RF votes, where each tree of the forest votes, based on f^n, if the voxel is representing a cancerous tissue. The summed up votes of the trees are a kind of prediction of how likely it is that the voxel is a cancerous tissue. The RF is applied on each voxel, and the result is stored in a prediction map. A threshold is applied on the map afterwards, defining how many votes are necessary to classify a voxel as cancerous. The default threshold of the RF algorithm is the majority vote, where more than half
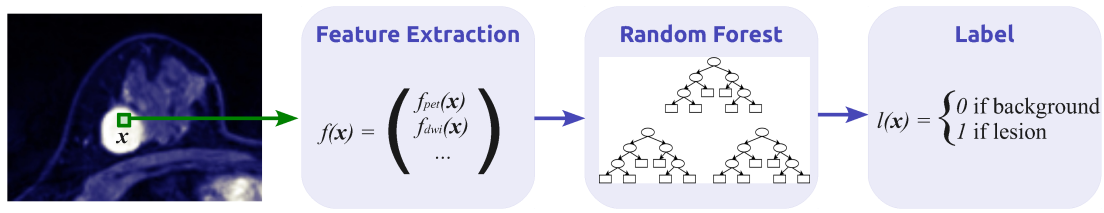
Figure 5.7: Illustration of the segmentation via RF classification for a given voxel at position $\mathbf{x}$. At first features are extracted from the image modalities (eg. f_{pet}, f_{dwi}). Then they are combined in a feature vector $f(\mathbf{x})$, which may contain all features or only a subset of the extracted features. A trained RF predicts the label $l(\mathbf{x})$ of the voxel depending on $f(\mathbf{x})$, wether it is a lesion or background.

of the trees need to vote for a class to assign it to a voxel. In this CAD pipeline an optimized threshold level is used, giving the highest Dice Similarity Coefficient in a Leave-One-Out Cross-Validation (LOOCV) process. More details to the threshold level are given in the Section 6.2 of the "Experiments" chapter. The final outcome of the segmentation step is a binary map l^n, in which every voxel within the breast is labeled as either cancerous or non-cancerous. Figure 5.8 illustrates the prediction map of a RF classification, and the threshold of the map.
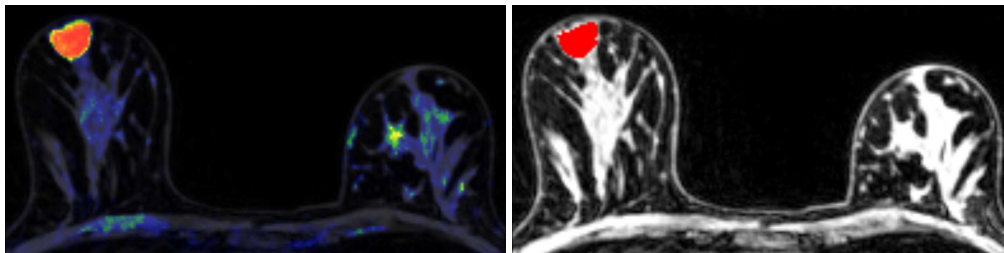
Figure 5.8: Left: Prediction map of a RF Classification. The color coding illustrates how many trees of the Random Forest voted for being cancerous. Blue means almost no tree voted and red means almost all trees voted for being cancerous tissue. Right: MRI overlayed with a binary map showing the segmented lesion (red). The binary image is gained by applying a threshold on the prediction map.

5.6 Lesion Classification Feature Extraction

The next step in the CAD pipeline is the classification of the lesion into being either benign or malign. Additional features are extracted from the images in order to get a prediction of the lesion class. Since the non-lesion area is not of interest in the classification step, only voxels within the lesion are considered. The segmentation map $l(\mathbf{x})$ obtained in the previous CAD step is used as mask for the feature extraction. Formally, the set of spatial coordinates within a lesion

is defined as $X_l = \{\mathbf{x}|l(\mathbf{x}) = 1\}$. The spatial voxel coordinates within the lesion are denoted as $\mathbf{x}_l \in X_l$. In the feature extraction step only spatial locations $\mathbf{x}_l$ are considered.

In contradistinction to the voxel based features $f(\mathbf{x})$ for the segmentation step, where features are extracted for each voxel, the lesion classification features f_l are extracted for the whole lesion. An obvious and simple extraction method would be averaging the data values within the lesion. Though, a high variability in the DCE-MRI contrast enhancement over the lesion is observed [36]. For instance necrotic areas have a much lower contrast enhancement than other regions. Such areas of low contrast-enhancement have an effect on the overall average value. Therefore, in the clinical practice a radiologist draws a small ROI over the region that appears to be the most enhancing region. The intensity values of the selected ROI are averaged and a kinetic curve is extracted from the averaged values [87]. The obtained kinetic curve can then be used for classification of the lesion [87]. The manual ROI selection suffers from a high intra- and interobserver variability in selecting a ROI within a lesion [110]. To overcome the variability Chen et al. [36] developed an method, which automatically selects a ROI and extract the kinetic curve by applying a FCM clustering on the DCE-MRI time-series. The cluster with the highest contrast enhancement was selected for classification. They denoted this cluster as the Characteristic Kinetic Curve (CKC) of the lesion. The clustering can be seen as a kind of dividing the lesion into several regions, from regions with a high contrast enhancement to regions with a low contrast enhancement. By picking the cluster with the high contrast enhancement, the ROI selection in the manual approach is imitated up to a certain degree, with the adavantage of the reproducibility of the automatic method.

The idea of Chen et al. of clustering DCE-MRI time-series is advanced in this thesis for DWI-ADC and PET data, which is a novel approach. The I_{adc} values within a lesion also have a high variability, partially due to the low resolution, partial volume effects and noise. In a manual approach a radiologist draws a ROI in a region with a low ADC value [24]. In this thesis I_{adc} is clustered and the cluster center with the lowest ADC value is chosen, imitating the manual ROI selection. In Section 6.3.1 of the "Experiments" chapter it is demonstrated that the clustering method produces discriminative features and that it is robust against outliers.

Formally, the intensity values from $I_{dce}(\mathbf{x}_l, i)$, $I_{adc}(\mathbf{x}_l)$, and $I_{pet}(\mathbf{x}_l)$ within a lesion are clustered using FCM. The obtained cluster centers μ_{dce}, μ_{adc}, and μ_{pet} are used as features for classification, resulting in one feature vector f_l per lesion.

In the following sections the FCM clustering is described in detail, as well as the clustering method of Chen et al. obtaining the CKC from the I_{dce} time-series, and the clustering of the I_{adc} and I_{pet} data using the FCM clustering method. At the end of this section a list of all features extracted for classification is given (Table 5.6).

5.6.1 K-Means and Fuzzy C-means Clustering of Features within Lesions

Data clustering is an unsupervised method of subdividing a sample data set I into c nonempty subsets (clusters) [22]. If each sample is assigned to one specific cluster, the clustering is termed as *hard* clustering. In a *fuzzy* clustering the strict assignment of a sample to one cluster subset is softened by using a degree of membership of a sample data point to each cluster.

The assignment to a cluster is guided by a similarity criterion, so that sample data values

within a cluster are more "similar" than data values of two distinct clusters. In a *centroid* based clustering the squared distance of a sample to its cluster center, which is the centroid (mean) of all samples within the cluster, is minimized over all samples. When defining a fixed cluster size c, the centroid based clustering can be formulated as an optimization problem: find the c cluster centers by assigning to each sample point a cluster so that the distances are minimized. Formally written: Given a set of sampled data or observations $\mathbf{I} = (I(\mathbf{x}_1), I(\mathbf{x}_2), ..., I(\mathbf{x}_n))$ with $x_j \in X_l$, the n samples are partitioned into c sets $\mathbf{S} = \{S_1, S_2, ..., S_c\}$ so that the sum of squares within the cluster is minimized:

$$\arg \min_s \sum_{i=1}^{c} \sum_{I(\mathbf{x}_j) \in S_i} \|I(\mathbf{x}_j) - \mu_i\|^2 \tag{5.11}$$

μ_i is the mean of all observations of the set S_i and represents the cluster center or centroid of the cluster i. The problem is NP-hard in regards of complexity [4]. One well known approximation algorithm, finding a local minimum of the solution, is the K-means algorithm by Lloyd [93]. It uses the fact that the optimal solution of the Eqn. (5.11) partitions the data into a centroidal *Voronoi* diagram [47], where the cluster centers $(\mu_1, \mu_2, ..., \mu_c)$ are the generating points of the diagram. The algorithm works as follows:

1. Start with an initial set of c cluster centers $\hat{\mu}_1^1, ..., \hat{\mu}_c^1$, either by placing the cluster centers randomly, or by using heuristics.

2. Partition the data according to the Voronoi diagram from the cluster centers by assigning each observation to the cluster with the closest cluster center. If two or more cluster centers are having the same distance, select one of them. Formally:

$$S_i^t = \{I(\mathbf{x}_p) : \|I(\mathbf{x}_p) - \hat{\mu}_i^t\| \leq \|I(\mathbf{x}_p) - \hat{\mu}_j^t\|; \ \forall 1 \leq j \leq c, 1 \leq p \leq n\} \tag{5.12}$$

3. Update the cluster center positions by calculating the mean for each cluster from the observations within the cluster:

$$\hat{\mu}_i^{t+1} = \frac{1}{|S_i^t|} \sum_{I(\mathbf{x}_j) \in S_i^t} I(\mathbf{x}_j) \tag{5.13}$$

4. Repeat steps 2-3 until the algorithm converges. Convergence is reached when there is no change in assignment of the observations to a cluster, or if the movement distance of the cluster centers is below a ε value.

The result of the K-Means clustering algorithm is the partition of the sample data in c clusters, and the corresponding cluster centers $\hat{\mu}_i$.

To explain the FCM algorithm, a $c \times n$ partitioning matrix $U = [u_{ij}]$ is introduced. u_{ij} represents the assignment of the observations $I(\mathbf{x}_j)$ to the cluster from set S_i. For a hard clustering

U is based on following constraints:

$$u_i(I(\mathbf{x}_j)) = u_{ij} = \begin{cases} 1; & I(\mathbf{x}_j) \in S_i \\ 0; & \text{otherwise} \end{cases} \tag{5.14a}$$

$$\sum_{j=1}^{n} u_{ij} > 0, \forall i \tag{5.14b}$$

$$\sum_{i=1}^{c} u_{ij} = 1, \forall j \tag{5.14c}$$

Eqn. (5.14a) specifies, that u_{ij} is 1, when $I(\mathbf{x}_j)$ is assigned to the cluster i. Eqn. (5.14b) restricts the clusters of having at least one observation assigned, whereas in Eqn. (5.14c) it is defined, that each observation is assigned to exact one cluster.

In a fuzzy clustering, based on an idea of Zadeh [179], the strict assignment of an observation to exactly one cluster is relaxed, by replacing the discrete binary assignment $u_{ij} \in \{0, 1\}$ in Eqn. (5.14a) with a "fuzzy" continuous function $u_{ij} \in [0, 1]$. The reformulated constraint for u_{ij} is:

$$0 \le u_{ij} \le 1; \text{ Eq. (5.14b); Eq. (5.14c)} \tag{5.15}$$

In a fuzzy clustering u_{ij} describes the degree of membership of the observation $I(\mathbf{x}_j)$ to cluster i in a range from 0 to 1. The hard clustering can be seen as an extreme case of the fuzzy clustering [48].

Based on the fuzzy clustering, Bezdek developed a FCM clustering algorithm [21,22], which works as follows: To introduce the degree of membership u_{ij}, the k-Means objective function (Eqn. (5.11)) is reformulated as:

$$J_b(U, V, Y) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}{}^b \|I(\mathbf{x}_j) - \mu_i\|^2 \tag{5.16}$$

where U is the partition matrix containing the degree of membership variables u_{ij}. V is the $c \times n$ matrix containing the cluster center vectors μ_i, and $b \ge 1$ is a weighting exponent, describing the amount of "unsharpness" (blur, defocus) of the memberships [22]. When setting b to 0 (no unsharpness), the formulation is equal to the K-means clustering formulation (Eqn. (5.11)) J_b is in general minimized, when the cluster centers μ_i are close to those observation points $I(\mathbf{x}_j)$ with a high degree of membership to cluster i.

Necessary conditions, where the objective function (5.16) is minimized under the constraints (5.15), can be derived using Lagrange multipliers [22]:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\|I(\mathbf{x}_j) - \mu_i\|^2}{\|I(\mathbf{x}_j) - \mu_k\|^2} \right)^{2/b-1}} \tag{5.17}$$

$$\mu_i = \frac{\sum_{j=1}^{n} u_{ij}^b I(\mathbf{x}_j)}{\sum_{j=1}^{n} u_{ij}{}^b} \tag{5.18}$$

The FCM clustering algorithm iteratively obtains the partition matrix U and the cluster center matrix V using Eq. (5.17) and Eq. (5.18) as follows:

1. Initialize the cluster center matrix V randomly or by using heuristics.

2. Update the cluster center matrix V using Eq. (5.17)

3. Update the partition matrix U using Eq. (5.18)

4. Repeat steps 2-3 until a convergence criterion ε is reached: $\|U^{t+1} - U^t\| < \varepsilon$

The result of the FCM clustering algorithm is a matrix V containing the cluster centers $(\mu_1, ..., \mu_k)$, as well as the partitioning matrix U containing the degrees of membership of each observation to each class.

5.6.2   Characteristic Kinetic Curve within Lesions

Within a lesion inhomogeneities in the contrast enhancement can be observed [36], resulting in a variability of time-signal curves within a lesion. Figure 5.9 illustrates the time-intensity curves of five randomly picked voxels, illustrating the variability of the time-signal curves. Chen et al. [36] proposed a method of categorizing these DCE-MRI time-signal curves into a number of prototypic curves by using a FCM clustering method and using the curve with the highest contrast enhancement rate, the CKC, for classification. They were able to show that using the CKC for classification gives better results than using curves obtained by averaging the intensities over the whole lesion.

This method is also used in the proposed CAD pipeline, and works as follows: Let $I_{dce}(\mathbf{x}_l)$ be a vector containing the intensity values of the DCE-MRI time-series at given position $\mathbf{x}_l$. Let furthermore $\mathbf{I}_l$ be the set of data samples within the lesion:

$$\mathbf{I}_l = \{I_{dce}(\mathbf{x}_1), I_{dce}(\mathbf{x}_2), ..., I_{dce}(\mathbf{x}_n)\} \tag{5.19}$$

where n is the number of voxels within a lesion. These samples $\mathbf{I}_l$ are clustered by a FCM algorithm. As cluster size c Chen et al. [36] empirically determined $c = \max(2, [n/80])$. They used $b = 2$ and the convergence criteria parameter $\varepsilon = 10^{-5}$. These parameters are adopted in the CAD pipeline, with an additional termination condition of 500 iterations as maximum number of iterations. The cluster centers V of the FCM are forming the prototypic curves of the lesion $(S_1, S_2, ..., S_c)$. The enhancement rate R_t for each prototypic curve is calculated in the similar way as in Equation (5.10). From these normalized curves the curve k with the maximum enhancement is chosen as the CKC:

$$k = \arg \max_{j=1,..,c} S_j(i) - S_{j0}; 1 \le i \le N \tag{5.20}$$

$S_j(i)$ is the signal intensity at time-point i for prototypic curve S_j and S_{j0} is the baseline intensity of the prototypic curve S_j. The baseline is obtained by calculating the mean intensity value over the first four time-points of the prototypic curve.

In Figure 5.9 the prototypic curves obtained from a lesion are plotted, and a comparison between the CKC and the curve received by averaging the intensities over the lesion is given.

The intensity values of the CKC are used as features for the classification. Formally:

$$f_{l\text{-}ckc} = (S_k(1), ..., S_k(N)) \tag{5.21}$$
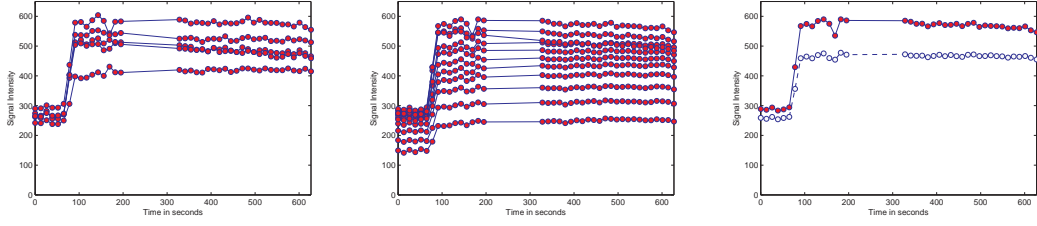
Figure 5.9: Left: Time-Intensity curves of five randomly picked voxels within the lesion. Middle: 11 prototypic curves obtained by FCM clustering. Right: CKC compared with the curve obtained by averaging the signal intensities (dotted curve).

In addition the values from the derivative of the CKC are calculated by the forward difference method (Eqn. (5.3)) and are also used as classification features:

$$f_{l\text{-}\Delta ckc} = (\frac{S_k(2) - S_k(1)}{t_2 - t_1}, ..., \frac{S_k(N) - S_k(N-1)}{t_N - t_{N-1}}) \tag{5.22}$$

5.6.2.1 Modeling of the Characteristic Kinetic Curve

To get the kinetic curve parameters (MaxEnh, TTP,...) the CKC is modeled by a regression curve, and properties from this curve are used as features for classification. Instead of fitting a third order polynomial, as proposed by Agner et al. [2], an asymmetric generalized logistic function [128] is used, multiplied with an exponential term to model the terminal slope. This regression curve has been developed by Grabner G. from the *MR Center of Excellence, Medical University of Vienna, Austria* and is published the first time in this thesis.

As a regression function an asymmetric generalized logistic function [128] multiplied with an exponential term to model the terminal slope is proposed:

$$r(t) = G * \left(1 - \frac{1}{\left(1 + (2^\alpha - 1) * \exp\left(\frac{1}{\tau} * (t - t_{1/2}) \right) \right)^{1/\alpha}} \right) * \exp\left(\beta * t^k \right) \tag{5.23}$$

G, α, τ and $t_{1/2}$ are the parameters of the asymmetric logistic function. G represents the scaling of the underlying sigmoid curve, α is the asymmetry parameter, τ is a constant that governs the steepness of the sigmoid curve, $t_{1/2}$ is the time where half of G is reached. $\exp\left(\beta * t^k \right)$ is the exponential term with the parameters β and k, where k defines the terminal slope of the curve, and β is a scaling factor of the exponential term. Figure 5.10 demonstrates the influence of the parameters on the shape of the regression curve. All model parameters are estimated by using a non-linear least squares regression method [142], which is provided by the Curve Fitting Toolbox of Mathwork's Matlab software.

The parameters α, τ, β, and k are used as features for the classification. G and $t_{1/2}$ are not used for further analysis, as the exponential function has an influence on G, and $t_{1/2}$ can vary due to time displacements, e.g. by a contrast media injection delay.

56
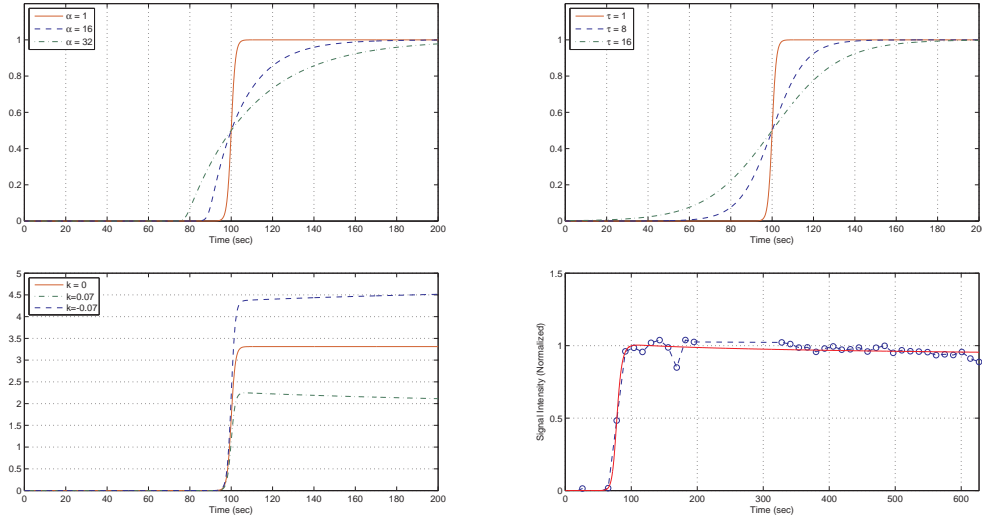
Figure 5.10: Illustration of the logistic model parameters, and the model fitted to a CKC. Top left: α defines the asymmetry of the logistic model; top right: with τ the steepness of the curve is defined. For both illustrations G has been set to 1, $t_{1/2}$ to 100, and k to 0. It can be seen that at time $t_{12} = 100$ the model function has a value of 0.5, which is the half of G, and the maximum of the function value is defined by the parameter G. Bottom left: The influence of k on the terminal slope is illustrated. Bottom right: The regression curve fitted to a given CKC.

From the fitted curve additional features are extracted, describing properties of the DCE-MRI time-intensity curve. The kinetic curve features are closely related to the curve features calculated for the segmentation (Section 5.4.1), but at this time the features are calculated from the regression curve, and not from the DCE-MRI intensity values. Following features are calculated: AuC, MaxEnh, THM, Time To Peak (TTP), as well as Maximum of DERivative (MDER), which is a replacement of the MaxSlope segmentation feature. For MaxSlope the maximum of the approximated derivative is used, whereas for the MDER the maximum of the analytical derivative $\frac{\delta r}{\delta t}$ of the regression function $r(t)$ is used, which is more precise than the approximation. All features are calculated for an 8 minute interval, beginning at the time-point t_0, at which the contrast enhancement starts. The starting point t_0 is determined by locating the time-point at which the intensity exceeds 0.5% of the maximum enhancement rate. To be tolerant towards delays in the contrast media injection, the interval has been limited to $t_0 + 8$ minutes; for this interval DCE-MRI data points are still available, even when the contrast media injection is delayed by two frames or 26 seconds. The formal definitions of the regression curve features are given in Table 5.5.

The feature vector $f_{l\text{-}curve} = (\alpha, \beta, \tau, k, AuC, MaxEnh, THM, TTP, MDER)$ represents the kinetic curve features used for classification.

Feature	Description	Calculation
AuC	Area under Curve	$$AuC = \int_{t_0}^{t_0+8\text{ min}} r(t)\,dt$$
MaxEnh	Maximum Enhancement	$$MaxEnh = \max(r(t)); \; t_0 \leq t \leq t_0 + 8\text{ min}$$
MDER	Maximum Derivation	$$MDER = \max\left(\frac{\delta r}{\delta t}\right)$$
THM	Time to Half Maximum	$$THM = \min(t) - t_0; \;\; \text{subject to } r(t) \geq MaxEnh/2$$
TTP	Time To Peak	$$TTP = \arg\max_t r(t) - t_0; \; t_0 \leq t \leq t_0 + 8\text{ min}$$

Table 5.5: Formal definition of the DCE-MRI kinetic curve features for classification. $r(t)$ is the logistic regression function with the time parameter t.

5.6.3 Clustering of DWI / PET data

In this thesis a FCM clustering is proposed to categorize I_{adc} and I_{pet}, with the aim of getting characteristic intensity values for a lesion, which are discriminative and robust against outliers. $I_{adc}(\mathbf{x}_l)$ values are collected in a set of data samples within the lesion:

$$\mathbf{I}_l = \{I_{adc}(\mathbf{x}_1), I_{adc}(\mathbf{x}_2), ..., I_{adc}(\mathbf{x}_n)\} \tag{5.24}$$

where $\mathbf{x}_i$ is the spatial coordinate within a lesion and n is the number of voxels. $\mathbf{I}$ is then clustered using a cluster size $c = 4$. The cluster center μ_{adc} with the lowest ADC value is chosen as the characteristic intensity value. Formally written:

$$f_{l\text{-}adc} = \min(\mu_{adc}^i), \; 1 \leq i \leq c \tag{5.25}$$

Choosing the minimum value as the characteristic value for the lesion is justified by the fact that malign lesions have in common a lower ADC value than benign lesions. In Section 6.3.1 of

Feature	Description	No. of Features
$f_{l\text{-}adc}$	Clustered DWI-ADC value	1
$f_{l\text{-}pet}$	Clustered PET value	1
$f_{l\text{-}ckc}$	Data values of the DCE-MRI CKC	40
$f_{l\text{-}\Delta ckc}$	Difference of values of the DCE-MRI CKC for two subsequent frames	39
$f_{l\text{-}curve}$	Parameter of the kinetic curve regression function: α, β, τ, k, AuC, MaxEnh, MDER, THM, TTP	9
	Sum:	90

Table 5.6: List of all features extracted for the classification. Features are grouped by 'Clustered Signal Intensity Features', and 'DCE-MRI Kinetic Curve Features'.

the "Experiments" chapter it is shown that using this method with a cluster size of 4 produces discriminative features for the classification. The I_{pet} values are clustered in a similar way, whereas in this case the cluster center μ_{pet} with the maximum value is chosen, resulting in a characteristic PET intensity value:

$$f_{l\text{-}pet} = \max(\mu_{pet}^i),\ 1 \le i \le c \tag{5.26}$$
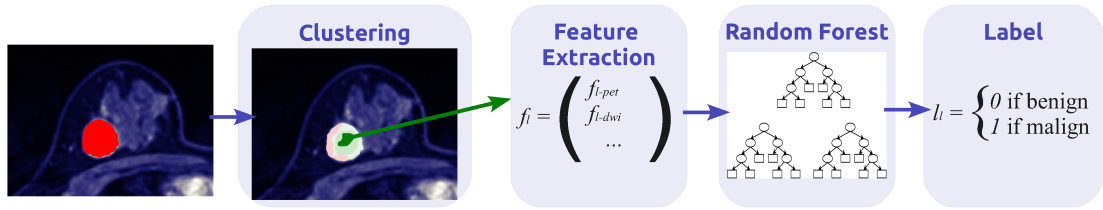
Figure 5.11: Illustration of the lesion classification process for a given lesion segmentation l. First the data within the lesion is clustered, and the cluster center with the highest enhancement for I_{dce} clustered data, the cluster center with the lowest value for I_{adc} clustered data, and the cluster center with the highest value for I_{pet} are chosen as features. Then they are combined in a feature vector f_l, which may contain all features or only a subset of the extracted features. A trained RF predicts the label l_l of the lesion depending on f_l, wether it is benign or malign.

5.7 Classification: benign vs. malign

The classification step is the final step in the CAD pipeline, where the segmented lesions are classified into malign and benign. A RF classifier is used for this task, where the decision is based on a subset f_l of the previously extracted features. The RF predicts the label l_l with $l_l = 0$ for benign lesions and $l_l = 1$ for malign lesions. Figure 5.11 illustrates the classification process. A performance evaluation of the extracted features can be found in Section 6.3.3 of the "Experiments" chapter.

RF lesion training   To learn a prediction model for a given feature subset f_l, the RF needs to be trained on this subset. Therefore a training feature set $f_{l\text{-}train}$ is evolved using the histopathological report as target label and f_l^i as training input data, where i is in this case the patient index. $f_{l\text{-}train}$ is then used to train a RF classifier.

RF Lesion classification   To classify a new lesion n, the feature vector f_l^n is extracted, containing the same features as used for the RF training. The RF uses this subset to predict the label l_l by letting each tree of the forest vote for a class. Let n_k be the number of votes for the class k. The probability p_k^n of the class k is then calculated by following equation: $p_k^n = \frac{n_k}{nTree}$. Finally, the lesion label l_l can be predicted by using the majority rule, where more than half of the votes of a RF prediction must be assigned to one class, to decide for this class. Other threshold values than the majority rule can be used to consider the misclassification costs. In this CAD pipeline an optimized threshold value is used, where the F value, which is a combination of sensitivity and specificity, is maximized in a LOOCV step.

The outcome of the CAD pipeline is a binary segmentation labeling $l(\mathbf{x})$, as well as a classification labeling l_l of the segmented lesion, whether it is benign or malign.

5.8 Discussion

The aim of this chapter was the description of the proposed CAD pipeline, from image acquisition, registration and segmentation to the classification of the lesion with respect to being benign or malign. The CAD pipeline is implemented as a process of six steps. In the first step the DCE-MRI and DWI images are acquired from a MRI scanner, and the PET images are obtained from a combined PET/CT scanner. As a second step the images are co-registered and scaled to a reference image, so that they are all aligned and all of them have the same spatial resolution. As registration method the SyN in combination with CC similarity metric for DWI/MRI, respectively PET/MRI registration, and MI similarity metrics for DCE-MRI registration have been proposed. In the third step the features are extracted for each voxel within the breast from the multimodal image dataset. A definition of the features, categorized into intensity based, texture based, and kinetic curve based features, have been given in this chapter. The segmentation of the lesion in step four is implemented as a binary classification process, where a RF machine learning algorithm decides, depending on the extracted voxel features, if a voxel is classified as cancerous tissue or normal tissue. As a result of this step a segmentation of the lesion is achieved. Based on the obtained segmentations features for the classification of the lesions are extracted. In contrast to the voxel-wise feature extraction for the segmentation step, classification features are extracted by clustering the values within the lesion and using the cluster centers as features. There are inhomogeneities within a lesion, partial volume effects at the border of the lesion and noise in the image modalities observed. The values within the lesions are categorized (clustered) and the mean value of the best discriminative category is used as feature. For DCE-MRI the cluster with the highest contrast enhancement, for DWI the cluster with the lowest ADC value and for PET the cluster with the highest PET intensity value are used. In addition to the intensity based features aasymmetric generalized logistic function is fitted to the kinetic curve obtained from the DCE-MRI clustering. Properties from this curve are used as features for classification. Finally, the classification in benign and malign lesions is done by a RF machine learning algorithm by using the clustered features in its prediction model. The outcome of the CAD pipeline is a binary segmentation labeling, as well as a classification labeling of the segmented lesion, whether it is benign or malign.

Experiments and Results

In this chapter the performance of the proposed CAD pipeline is evaluated. The aim, beside of the CAD pipeline evaluation, is the validation of the hypothesis that a multimodal medical imaging approach improves the performance of a breast lesion CAD system in comparison to a single modality CAD system. Thus, one main point of the experiments relies on the exploration of the DWI and PET modalities in combination with DCE-MRI.

Outline   This chapter is structured into three main sections, corresponding to the three main parts of the CAD pipeline: registration, segmentation, and lesion classification.

1. **Registration.** In Section 6.1 transformation models and similarity metrics suitable for multimodal breast image registration are evaluated. The quality of the registration is verified for CT to MRI, DWI to MRI and DCE-MRI registration.

2. **Segmentation.** In Section 6.2, the segmentation is evaluated. First, the proposed segmentation features are analyzed, using the variable importance functionality of the RF. Features are evaluated with regard to the relevant information they provide to the voxelwise segmentation. The performance of the segmentation itself is examined, using several combinations of features. The influence on the accuracy of the segmentation for each modality and each category of features is explored.

3. **Classification.** The final section of this chapter (Section 6.3) addresses the performance evaluation of the classification of breast lesions into benign and malign. The outline of this set of experiments is similar to the experiments from the segmentation section. First the classification features are analyzed using the variable importance score of the RF. Then the performance and accuracy of the classification is evaluated for several groups of features. Also the influence of the three modalities on the classification performance is explored.

Each section is closed with a discussion, summarizing and reflecting the results of the experiments.

6.1 Registration

In this section the registration performance using three transformation models and similarity metrics is evaluated. Registrations from CT, DCE-MRI and DWI to the MRI modality are covered. The aim of these experiments is to find the best performing transformation model and similarity metric for the given registration task.

As a similarity metric CC and MI are evaluated for CT to MRI, DCE-MRI time-series to MRI and DWI to MRI. For DCE-MRI MSQ metric is also used. This metric has been chosen, because it is less computational demanding than CC and MI, as reported by Avants et al. [14].

The types of transformation models used for the registration are Elastic (EL) [57], Diffeomorphic Demons (DD) [156], Symmetric Geodesic Normalization (SyGN), and Greedy Symmetric Normalization (GrSyN) [14]. The SyGN transformation model is used only for CT to MRI registration, and has been skipped for the other modalities in the subsequent tests, due to the high computational effort needed for the registration. The registration process revealed that the time needed for registration is one magnitude higher than for other transformation models. On an Intel Xeon CPU X5450 3.0 GHz the registration time for one breast image was between 10 and 20 hours, whereas for other modalities the maximum registration time was 3 hours.

6.1.1 Experimental Setup

As a first step an affine registration from CT, DCE-MRI and DWI to the MRI modality is applied, to globally align the breast images. As similarity metric for the affine registration the mutual information is used, since it is known to provide robust results in a rigid multimodal registration process [12]. Based on this affine registration non-rigid models are evaluated. Both registration types, affine and non-rigid, are computed on a multiresolution scale based on a 3 level Gaussian pyramid. The initial registration begins with a quarter of the original resolution, the second turn uses half of the resolution and the final turn is performed with the full resolution. The maximum number of iterations is limited to 500, 300 and 150 in respect to the resolution level, after initial experiments. As regularization a Gaussian filter with a σ^2 of 3 is used. Both, the number of resolution levels and the value for σ^2 have been suggested by Avants et al. [14] as standard registration parameters for MRI images with 1 mm^3 resolution.

To evaluate the performance of DCE-MRI time-series registration, $I_{dce\text{-}peak}$ is exemplarily chosen for registration in the following experiments.

6.1.2 Validation Measures

MI is used to evaluate the quality of the registration. It is calculated using the voxels within the breast ($\mathbf{x} \in M$) from the registered volume and the MR volume, whereas voxels from the surrounding air are masked out and not considered for the calculation. A higher MI value indicates a better registration. In addition to the MI metric a Dice Similarity Coefficient (DSC) [45] is calculated from a threshold basic segmentation. The DSC is defined as:

$$DSC(A, B) = \frac{2 \times |A \cap B|}{|A| + |B|} \tag{6.1}$$

Figure 6.1: Threshold CT, DCE-MRI and DWI (from left to right), on which the transformation has been applied to. These threshold images are used for the DSC calculation. The transformation model in this example is GrSyN and metric is CC.

A are the foreground labels of the first image and B are the foreground labels of the second image. The range of $DSC(A, B)$ is between 0 and 1, where 0 indicates no overlap and a value of 1 indicates a perfect match.

The segmentation is obtained by applying a threshold on the image. Otsu's method [113] is used to determine the optimal threshold level, resulting in a rough binary segmentation of the breast tissue. The transformation yielded from the registration process is applied to the threshold images, and the DSC is calculated from the transformed threshold images.

The CT threshold image is compared with the binary mask M of the MR breast image instead of the MR threshold image. This is necessary, because at the CT threshold image the whole breast, including the fat-tissue is declared as foreground, whereas in the MRI threshold image the fat-tissue parts are declared as background, due to the low contrast of fat-tissue in fat-suppressed MRI. In this case the DSC measures the similarity of the registered CT breast shape to the MRI breast shape, whereas the accuracy of the registration within the breast is measured by MI.

For DCE-MRI / MRI and DWI / MRI threshold images the DSC coefficient has a slightly different meaning. The MR images are recorded using fat-suppression and also DWI values are low for fat-tissue. Thresholding this kind of images set the low-intensity fat tissue as background and the other tissue as foreground voxels. The DSC is therefore a measurement of the overlap of non-fat tissue in DCE-MRI and DWI. Image 6.1 shows the transformed threshold images for one patient.

The registration and the calculation of MI and DSC are done for 8 patient images and the mean and standard deviation of MI and DSC are calculated.

6.1.3 Results

CT to MRI Registration   The evaluation results of the CT to MRI registration are summarized in Table 6.1. Transformation models using MI as the similarity metric perform worse than their CC counterpart. A lower DSC is also observed for MI similarity metric, which indicates that the shape of the breast is not fitted as well as when using the CC metric. SyGN in combination with CC gives the best results, followed by GrSyN.

In Figure 6.2 the registration process is visualized. It can be seen that in addition to the rigid registration a deformation registration is necessary to match the images. Figure 6.3 shows the registration results for the given similarity metrics and transformation models. Especially at the

Sim. Metric	Transf. Model	Mean MI	Std MI	Mean DSC	Std DSC
Mutual Information (MI)	DD	0.5424	0.1279	0.8450	0.0473
	EL	0.5716	0.1029	0.7470	0.0712
	GrSyN	0.6546	0.1728	0.8607	0.0463
	SyGN	0.6183	0.1460	0.7685	0.0668
Cross-Correlation (CC)	DD	0.6161	0.1228	0.8885	0.0452
	EL	0.6646	0.1187	0.8890	0.0480
	GrSyN	0.7125	0.1147	0.8930	0.0470
	SyGN	**0.7527**	0.0949	**0.8944**	0.0459

Table 6.1: MI and DSC calculated using different types of similarity metrics and transformation models for registration of CT and MRI modalities. DD - Diffeomorphic Demons style exponential mapping, EL - Elastic transformation, GrSyN - Greedy Symmetric Normalization, SyGN - Symmetric Geodesic Normalization
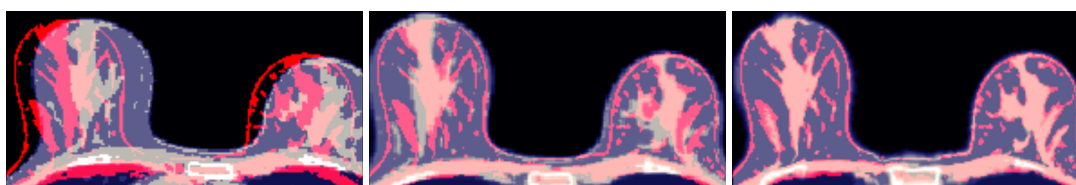
Figure 6.2: Registration steps and results from the CT modality to the MRI modality. From left to right: Unregistered images, affine registration, GrSyN registration with CC similarity metric. The images in this figure are thresholded for illustration purpose. The thresholded MRI image is red colored, the CT image is blue-gray colored.

regions next to the Pectoralis muscles a flawed registration can be observed for MI similarity metrics. In this example it can be also seen that the shape is not well fitted using EL in combination with MI. This behavior has been observed for registration results of several datasets using MI as similarity metric. These kind of flaws result results in a lower DSC score for the MI similarity metric, due to the reduced shape overlap.

DCE-MRI to MRI Registration   In Table 6.2 the results of the DCE-MRI to MRI registrations are summarized. In this case the MI similarity metric performs better than the CC variant, and MSQ has the worst performance of all three. The contrast enhancement in the DCE-MRI image and the resulting difference in the image intensities have a negative impact on the MSQ metric. In this experiment the SyN transformation model also outperforms the other transformation models.In Figure 6.4 the registration is visualized. It can be seen that there is only a small offset of the unregistered DCE-MRI image to the MRI image, which is a consequence of small patient movements and breathing during the image acquisitions.

DWI to MRI Registration   The registration evaluation from DWI to MRI is summarized in Table 6.3. The CC similarity metric provides better results than the MI metric. Also in this
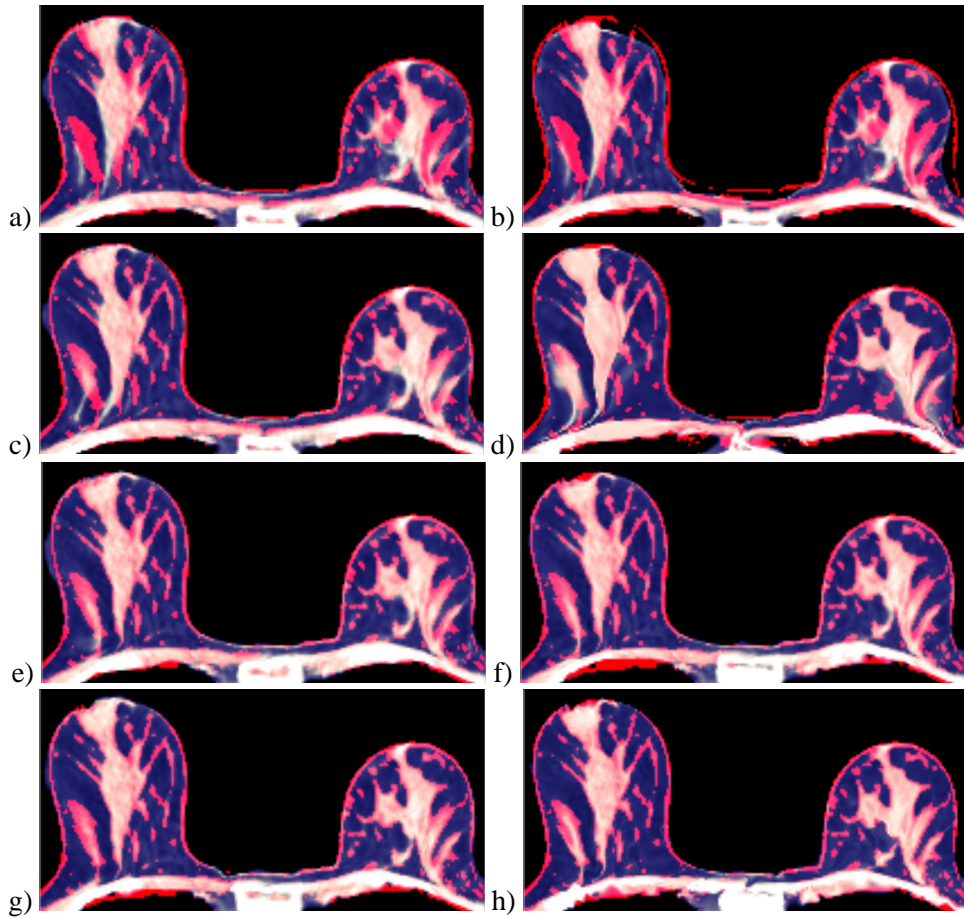
Figure 6.3: Registration results from the CT modality to the MRI modality for one patient. For a better orientation and comparison, a threshold MRI has been overlayed in red colors. The blue-gray colored image is the registered CT image. Similarity metric for the first 4 images a)-d) is MI, for the second 4 images e)-h) it is CC. The registration transformations are: DD for a) respectively e), EL for b) respectively f), GrSyN for c) respectively g), and SyGN for d) respectively f).

case the GrSyN transformation surpasses the other models. Figure 6.5 shows the registration result for one patient.

6.1.4 Discussion

The aim of this experiment was the evaluation of similarity metrics and transformation models in the context of multimodal breast image registration. The registration was performed on CT to MRI, DCE-MRI to MRI, and DWI to MRI, using images from 8 patients. The performance was evaluated using MI, calculated from the voxels within the breast of the transformed images and the reference MRI image. In addition, the DSC has been computed from binary threshold

Sim. Metric	Transf. Model	Mean MI	Std MI	Mean DSC	Std DSC
Mutual Information (MI)	DD	1.5875	0.2406	0.7826	0.0512
	EL	1.6154	0.2393	0.7758	0.0537
	GrSyN	**1.6980**	0.2200	**0.8067**	0.0418
Cross-Correlation (CC)	DD	1.4882	0.2052	0.7786	0.0464
	EL	1.4975	0.2029	0.7753	0.0474
	GrSyN	1.5335	0.1968	0.7752	0.0472
Mean Square Difference (MSQ)	DD	1.0436	0.1750	0.6735	0.0591
	EL	0.8779	0.2069	0.6070	0.0641
	GrSyN	1.1261	0.1800	0.6068	0.0511

Table 6.2: DSC and MI score computed from registration of DCE-MRI to MRI using several similarity metrics (MI,CC,MSQ) and transformation models (DD,EL,GrSyN).

Figure 6.4: Unregistered (left) and registered (right) DCE-MRI image(white) with a MRI image(red) overlayed for orientation. Threshold of images are used for visualization purpose only.

Sim. Metric	Transf. Model	Mean MI	Std MI	Mean DSC	Std DSC
Mutual Information (MI)	DD	0.2919	0.1072	0.1490	0.0694
	EL	0.3108	0.1092	0.0808	0.0491
	GrSyN	0.3447	0.1232	0.2119	0.0847
Cross-Correlation (CC)	DD	0.3256	0.1027	0.2112	0.1020
	EL	0.3504	0.0953	0.2099	0.1034
	GrSyN	**0.4154**	0.0892	**0.2135**	0.1018

Table 6.3: DSC and MI score computed from DWI to MRI registration using several similarity metrics (MI,CC,MSQ) and transformation models (DD,EL,GrSyN).

images, to measure the similarity of the shape of the two breast images.

The LDDMM approach, SyN, provided the best results. In the first test, the registration of CT and MRI images, both SyN variants, greedy and geodesic, have been evaluated. The geodesic variant surpassed the greedy variant. However, the high computational demand of the geodesic variant of SyN limits its usefulness.

For registration of CT/MRI and DWI/MRI images the CC similarity metric outperforms the MI metric. The reason is that the linear intensity relationship of the tissues in the breast,
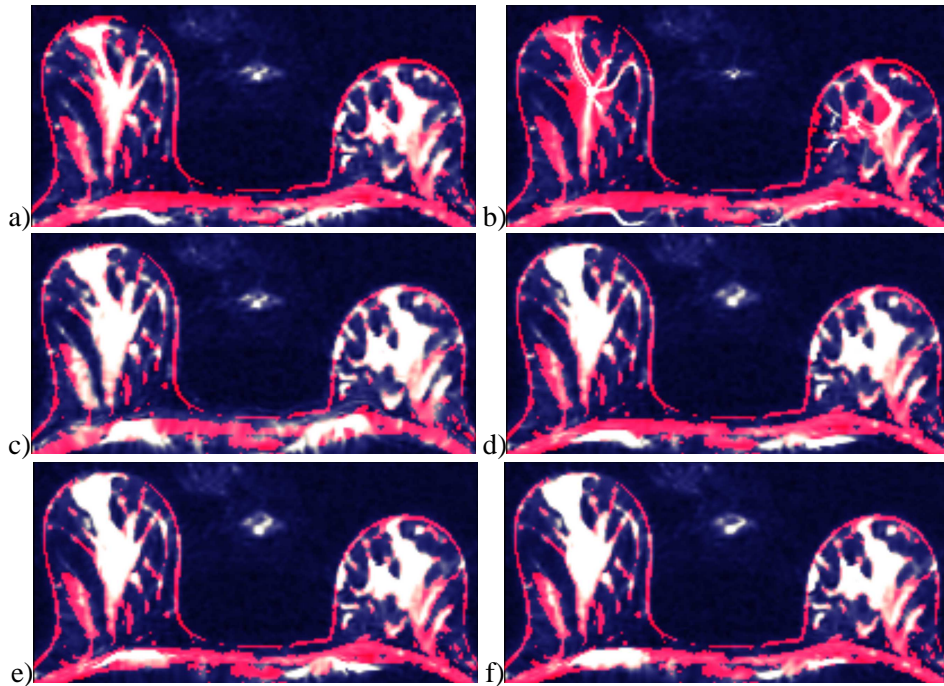
Figure 6.5: Registration results of DWI modality for one patient. Similarity metric for images a)-c) are MI, for d)-f) it is CC. The registration transformation models are: DD for a) and d), EL for b) and e), GrSyN for c) and f). The MRI modality is red colored, the DWI modality is blue-gray colored.

like the fat tissue and the lobules, is a good guidance in the registration process. The reduced performance of the MI metric may be deduced from the fact that the performance of MI on a local scale is reduced when intensity inhomogeneities exists in the image [12,143] (i.e. MRI bias fields). Such bias fields are observed in the MRI breast images. The DCE-MRI/MRI registration performance is not influenced, since both images are acquired in the same sequence, and share similar bias fields. The higher performance of MI for MRI images is in coincidence with this hypothesis. However, this hypothesis is not investigated in detail in this thesis, and may be explored in detail in further studies.

The following experiments are based on the results of the registration experiments. As prerequisite to the following segmentation and classification experiments, all patient images are registered using GrSyN and CC for CT to MRI and DWI to MRI registration, as well as MI for DCE-MRI to MRI registration.

6.2 Segmentation

In this section the segmentation performance and the influence of feature groups on the segmentation are evaluated. The following experiments aim at on the one hand finding the best feature combination for segmentation, and on the other hand elaborating the advantage of a multimodal approach to a single DCE-MRI modality approach in the context of lesion localization and segmentation.

The outline of this section is as follows:

- **Multimodal feature selection via Gini importance**: First, the contributions of the features are explored using the Gini importance measure of the RF classifier. Based on the importance measure a feature selection is performed and evaluated.

- **Gini importance of the DCE-MRI time-series**: The Gini importance of DCE-MRI time-series features (intensity, texture) are analyzed to evaluate which part of the DCE-MRI time-series carries information with regard to lesion segmentation.

- **Segmentation performance for feature subsets**: The segmentation performance is evaluated using subsets of features. It is analyzed, which feature set improves the segmentation performance, and the benefit of using multimodal features with regard to lesion segmentation.

- **Qualitative Analysis of the Segmentation Results**: Finally, the segmentation results are plotted and discussed.

To recall the features used in the segmentation and in the following experiments, one may be referred to Table 5.4 in the previous chapter.

Experimental Setup For the segmentation of the lesion the data from 16 patients are used. Only voxels from within the breast are considered, the surrounding air is masked out and not used in the training, prediction, and performance calculation by using the masking M. For each patient the lesions have been voxel-wise annotated by a radiologist (L^i). A RF is trained on the feature set $f_{train}(\mathbf{x})$ of lesion and non-lesion voxels with the target labels $L(\mathbf{x})$. The trained RF is then used for prediction of voxels being in the first or second class.

Since a breast image contains 1 Mio. voxels and more, and since there is a massive imbalance in the size of samples in both classes (> 1 Mio. non-lesion, a few thousand lesion voxels), *random undersampling* [70] is performed in order to get a reasonable size of training data and to balance the training data. 1000 samples from the annotated lesion and 5000 samples from the other breast tissue are drawn randomly per patient. The small imbalance between background and foreground samples has been chosen to consider the higher variability of the background. The sampling size has been empirically determined. It has to be mentioned that random undersampling may result in an information loss, causing the classifier of missing relevant concepts in the data [70]. Therefore the number of samples needs to have a certain size to be statistically meaningful. However, using more advanced subsampling methods may reduce the needed sample size and increase the performance by reducing the information loss in the sampling process. He and Garcia describe this problematic of undersampling in detail in [70].

The number of trees (*ntree*) in the RF training is set to two times of the number of features used, with a minimum size of 200. To take care of the 5:1 imbalance of the non-lesion and lesion samples the class weight parameter of the RF has been also set to 5:1 (*classwt*). Chaoe Chen et al. describe the effect of weighted RF in [33]. And as a last parameter for the RF the number of input variables tried at each split, (*mtry*), is set to $\lfloor \sqrt{\# \text{ features}} \rfloor$ as suggested by Breiman [28].

This setup is used for all following segmentation experiments, unless otherwise stated.

Validation measures   The validation of the segmentation performance is performed by using Leave-One-Out Cross-Validation (LOOCV). The dataset from one patient is used for validation and the sampled dataset from the other patients are used for training. This is repeated so that each patient is validated once.

The quality of the segmentation is measured by comparing the predicted segmentation $l(\mathbf{x})$ with the manual annotated data $L(\mathbf{x})$ using DSC as similarity metric. Instead of using the majority vote in the RF prediction, the threshold level for the RF votes is set in a way that the mean DSC score is maximized. Boxplots [99] are used to visualize the statistical summaries, and in particular the variance in the DSC score among the patients.

6.2.1   Multimodal Feature Selection via Gini Importance

In the following experiments the segmentation features are explored using the Gini importance obtained from a trained RF. The Gini importance measure is used on the one hand to get an insight into the relevance of the features for segmentation, and on the other hand it is used for feature selection by using the Gini importance measure to rank the features, and to drop features with a low score. An interesting detail is that the RF algorithm does not have any a-priori information about these features, and the relevance of the features is solely determined in the training step.

The RF feature selection is also compared with another feature ranking method proposed by Peng et al. [115]. Their method is called *minimum-Redundancy-Maximum-Relevance (mRMR)* feature selection. It provides a ranking of the features based on Mutual Information, where the relevance of features and the redundancy between features are considered and rated. The relevance $D(S, c)$ of the feature set S is calculated by averaging the MI among all features f_i and the target class c (Equation (6.2)). The redundancy $R(S)$ is measured by calculating the mean of MI between two features f_i and f_j (Equation (6.3)). Maximizing the relevance and minimizing the redundancy can be combined by calculating the difference of $D(S, c)$ and $R(S)$. This leads to the definition of mRMR as described in Equation (6.4).

$$D(S, c) = \frac{1}{|S|} \sum_{f_i \in S} MI(f_i, c) \tag{6.2}$$

$$R(S) = \frac{1}{|S|^2} \sum_{f_i, f_j \in S} MI(f_i, f_j) \tag{6.3}$$

$$\max_S \Big[\frac{1}{|S|} \sum_{f_i \in S} MI(f_i, c) - \frac{1}{|S|^2} \sum_{f_i, f_j \in S} MI(f_i, f_j) \Big]. \tag{6.4}$$

71
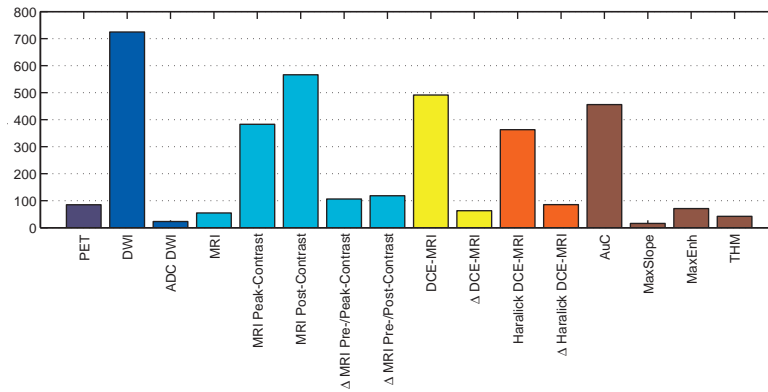
Figure 6.6: Gini importance. The blue bars are features from PET and DWI, cyan bars are high-resolution MRI and DCE-MRI features, yellow are DCE-MRI time-series features, orange are Haralick texture features, and red are DCE-MRI kinetic curve features. For the DCE-MRI features and Haralick texture features only the value with the highest Gini importance measure of the whole time-series is plotted.

Experimental Setup   To get the Gini importance and subsequently the ranking of the features, a training of the RF is performed using all features and all of the sampled training dataset.

For the mRMR feature selection, and in particular for the MI calculation, the data values are quantized in 5, respectively 32 bins. In addition, the MI is calculated from the continuous values using a Parzens windowing. Peng et al. [115] suggest a maximum of 5 states per feature variable, therefore 5 bins have been chosen. However, experiments revealed that there is a fluctuation in the mRMR ranking with varying bin size. The ranking became stable with a size of 15 bins and more. Therefore, 32 bins have been chosen as additional bin size for this experiment.

Validation measures   To evaluate the quality of the ranking the features are successively reduced using the yielded rankings. A RF is trained on the reduced feature set. The performance of the feature subset is evaluated as described at the beginning of this section by using LOOCV and DSC as similarity measure. The performance value is obtained by calculating the mean of the DSC scores for each patient. The evaluation is performed for a feature subset size of 200, 100, 50, 25, 12, and 6 to 1 using the ranking obtained from the Gini importance, respectively the mRMR.

Results   In Figure 6.6 the Gini importance score is summarized. The 12 highest ranked features obtained from the Gini index, respectively the mRMR are summarized in Table 6.4. The color coding used in the table refers to the colors used in the summary Figure 6.6.

The performance of the feature selection is plotted in Figure 6.7. The mean DSC is quite stable up to a reduction to the 5 top-ranked features, even that RF and the mRMR algorithm chose a different set of features. Therefore, a higher amount of features does neither increase nor decrease the performance of the segmentation. The RF algorithm used for training and
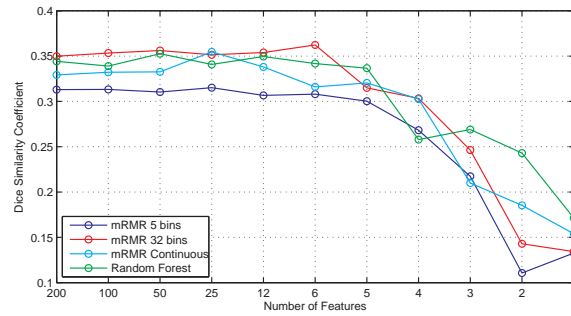
Figure 6.7: Mean DSC score using the reduced feature set. Subset selection is done by using the RF Gini importance value and the mRMR method. Discretization of the data for the mRMR method is done by quantification into 5 and 32 bins.

prediction inherently includes a feature selection, preferring "important" features. Going below 5 features removes features which are essential for prediction, resulting in a drop in the score.

The performance of RF feature selection is comparable to mRMR, whereas mRMR with 32 bins performs best, up to a reduction to 6 features. Using the features of the mRMR with 5 bins performs worst.

There is a drop in the DSC of the 32 bin mRMR observed, when reducing from 6 to 5 features. In this case the PET feature is dropped, which has quite valuable information needed for a better segmentation, at least for this combination of features. There is also a drop from 5 to 4 features of the RF feature selection. In this case the AuC feature is dropped. In all the feature selection algorithms the DWI feature is ranked in top positions, except for the 5 bins mRMR feature selection. Removing this feature from the set causes a drop of the DSC (Feature #4 in 32 bin mRMR and feature #3 in continuous mRMR). This indicates that the DWI feature also contains valuable information for the segmentation process.

The DCE-MRI Peak and Post-Contrast features and their derivatives are also ranked high in all of the feature selection methods. It has to be noted that the annotations of the lesions were drawn on the registered DCE-MRI Peak and Post-Contrast images. Also the DCE-MRI Peak and Post-Contrast images have the highest resolution of all modalities, therefore partial volume effects at the border of the annotated lesion are smaller, resulting in a more precise segmentation.

One interesting point of the experiment result is, that even when a complete different set of features are selected, the performance is quite similar. This shows, that many features share redundant information, and therefore can be combined in different ways. It is also interesting to see, that the RF also do a comparable feature selection to mRMR, even though there is no mechanism in the RF handling the redundancy of the information in features. The mRMR algorithm measures the redundancy of the features via Mutual Information.

	Random Forest	mRMR Continuous
1	DWI	DCE-MRI Peak-Contrast
2	DCE-MRI Post-Contrast	Haralick #6 DCE-MRI frame #15
3	DCE-MRI frame #22	Δ Haralick #7 DCE-MRI frame #4
4	Haralick #6 MRI Post-Contrast	DWI
5	Curve Feature: AUC	Δ MRI Pre-/Post-Contrast
6	DCE-MRI frame #30	Haralick #11 MRI Post-Contrast
7	DCE-MRI frame #24	Haralick #5 DCE-MRI frame #6
8	DCE-MRI frame #20	Haralick #6 MRI Post-Contrast
9	DCE-MRI frame #21	Curve Feature: Maximum Enhancement
10	MRI Peak-Contrast	Haralick #11 MRI Peak-Contrast
11	Haralick #6 DCE-MRI frame #15	DCE-MRI frame #2
12	DCE-MRI frame #27	Δ MRI Pre-/Peak-Contrast

	mRMR 32 bins	mRMR 5 bins
1	DCE-MRI Post-Contrast	DCE-MRI frame #38
2	Haralick #6 DCE-MRI Frame #4	Haralick #6 MRI Pre-Contrast
3	DWI	Δ Haralick #6 DCE-MRI frame #4
4	Δ Haralick #10 DCE-MRI frame #4	PET
5	Δ DCE-MRI frame #7	Δ MRI Pre-/Post-Contrast
6	PET	Haralick #6 DCE-MRI frame #19
7	Δ Haralick #6 DCE-MRI frame #4	MRI Peak-Contrast
8	Δ MRI Pre-/Peak-Contrast	Δ Haralick #9 DCE-MRI frame #4
9	Δ Haralick #7 DCE-MRI frame #13	Δ DCE-MRI frame #7
10	Haralick #11 MRI Post-Contrast	Δ Haralick #11 DCE-MRI frame #4
11	DCE-MRI Frame #2	Haralick #6 MRI Post-Contrast
12	Δ Haralick #4 DCE-MRI frame #3	MRI Post-Contrast

Table 6.4: The twelve top-ranked features using RF Gini importance measure and mRMR ranking. The color of the cells are similar to the colors used in Figure 6.6. PET and DWI features are blue, high resolution MRI features are cyan, DCE-MRI features are yellow, Haralick features are orange, and DCE-MRI kinetic curve features are brown.
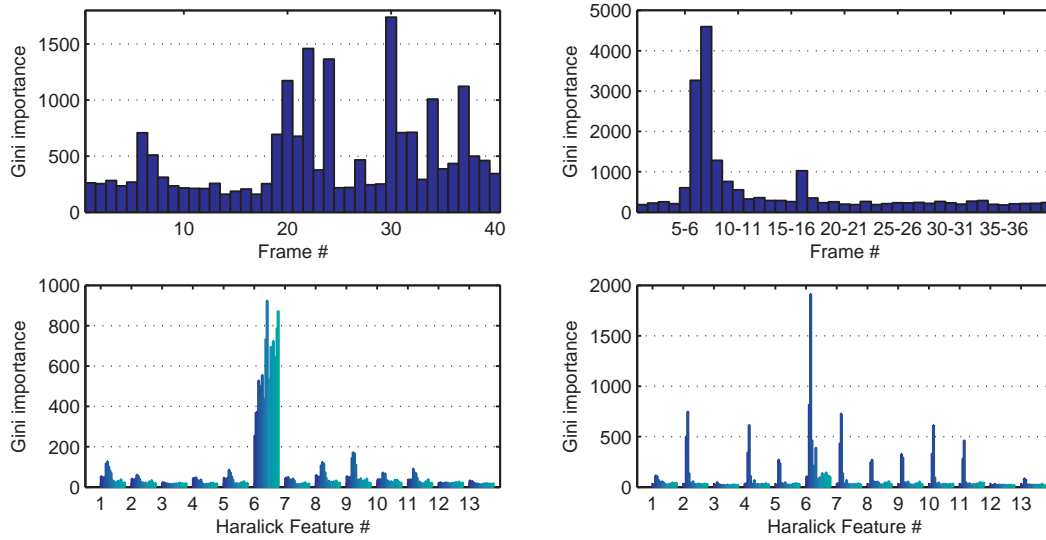
Figure 6.8: Gini importance of a RF trained on the DCE-MRI time-series. The upper left diagram shows the importance of the features f_{dce}. The upper right diagram shows the importance of the features $f_{\Delta dce}$. On the lower left side the importance of the Haralick features $f_{tex\text{-}dce}$ is plotted, grouped by the specific Haralick feature number. Each subhistogram represents the time-series for the corresponding Haralick feature. The lower right diagram shows the importance of $f_{\Delta tex\text{-}dce}$, also grouped by the Haralick statistics.

6.2.2   Gini Importance of the DCE-MRI Time-series

In this experiment it is evaluated which part of the DCE-MRI time-series carries information with regard to lesion segmentation. The DCE-MRI sequence contains forty images, from which 40 DCE-MRI, 39 Δ DCE-MRI, as well as 13×20 Haralick texture features and 13×19 Δ Haralick texture features are extracted, resulting in 507 DCE-MRI based features. The following experiment takes a closer look on these DCE-MRI features. It is analyzed, which features of the DCE-MRI time series, and which Haralick feature contains valuable information for the segmentation process. As in the previous experiment the Gini importance yielded from RF training is used for exploration and interpretation of the DCE-MRI features.

Experimental Setup   A RF with 400 trees is trained using the sampled training dataset, the annotated lesions, and an alternating feature subset containing the DCE-MRI time series features f_{dce}, $f_{\Delta dce}$, the Haralick DCE-MRI texture features $f_{tex\text{-}dce}$, respectively $f_{\Delta tex\text{-}dce}$.

Validation Measures   From each trained RF the Gini importance measure is retrieved. Each Gini importance gives a measure of importance of the DCE-MRI and Haralick texture features, respectively their derivatives.

Results   The Gini importance for each group is plotted in Figure 6.8. When using f_{dce} for RF training, it can be seen that the algorithm prefers frames from the middle and the last thirds of the time-series, which is the late postcontrast phase. In this phase the contrast or signal intensity of the lesion is higher than for non-lesion tissue, providing important discriminative features for the RF classifier. A small peak at frame number 6 and 7 can also be observed. At this time-point the contrast uptake takes place.

In the Gini importance measures of $f_{\Delta dce}$ a significant peak at frame 6-7 and 7-8 can be observed. This is also the time point of the contrast uptake, where the intensity at the lesion region increases fast. There is a second peak at frame number 16-17. Between these two frames there is a two-minute gap in the sequence, where the high-resolution $I_{dce\text{-}peak}$ is recorded. In this case the difference between these two time points is accumulated over two minutes, whereas at the other time points there is only an accumulation of 13 seconds.

For $f_{tex\text{-}dce}$ it can be observed that the RF prefers the Haralick feature #6, which is the "sum average". A high value corresponds to a region with a high intensity value; respectively, a region with low intensity value has a low sum average. This observation covers well with the Gini importance of f_{dce} , where features with a high intensity value are chosen at the lesion region during the DCE-MRI sequence.

And in the last diagram of the Figure 6.8, the $f_{\Delta tex\text{-}dce}$ features, it can be seen that there is a peak at frame 3-4 and 4-5 for almost all features with the highest Gini importance for the "sum-average" feature (Feature #6). Since the Haralick features are only calculated for every second frame, these frames correspond to the frame 6-8 in the original DCE-MRI time series. This is the time-point of the contrast uptake.

6.2.3   Segmentation Performance for Feature Subsets

In the first experiment of this section the RF has been used to select a feature subset from *all* features. In the following experiments the RF is trained on a limited set of features. The sets of features are grouped by the modality (MRI, DWI, PET) and the category (intensity features, texture features, kinetic curve features) for the MRI modality. Using groups of features for training and prediction enables a comparison of the segmentation performance, when using a single modality, respectively a multimodal approach. Also a comparison of intensity, texture, and kinetic curve features is possible in such a setup. By restricting the feature set in the RF training to a limited group of modalities and categories a deep insight in each of the categories is given. The benefit of each modality and each category on the segmentation performance is evaluated in the following experiment.

Experimental Setup   The features are divided into following groups: f_{dwi}, f_{pet}, $f_{dce\text{-}pre}$, f_{dce}, $f_{\Delta dce}$, $f_{tex\text{-}dce}$, $f_{\Delta tex\text{-}dce}$, and f_{kin} (AuC, THM, MaxEnh, and MaxSlope). In the subsequent experiment single groups of features, respectively combinations of them are used for RF training and prediction.

Validation Measures   The segmentation performance is evaluated in a LOOCV process, calculating the DSC score for each patient. From these DSC scores the mean and standard deviation

Feature Type	Mean DSC	Std. DSC
f_{dce}	**0.3952**	0.2920
f_{kin}	0.3297	0.2809
$f_{\Delta dce}$	0.3694	0.3385
$f_{tex\text{-}dce}$	0.3131	0.2195
$f_{\Delta tex\text{-}dce}$	0.3158	0.2715

Table 6.5: DSC statistics for DCE-MRI feature subsets.

is computed and the results are visualized using boxplots. In addition, the DSC score is plotted as bar for each patient, enabling a visual comparison of the feature groups on a per-patient level.

6.2.3.1 Results

To get a better overview the results are grouped by the modality and category. First the results of the DCE-MRI modality are presented, focusing on the intensity, texture and kinetic curve features, and combinations thereof. The second part focuses on the modalities MRI, PET, and DWI, where the performance of the individual modalities and combinations of them is evaluated. It is also analyzed, if a combination of PET, DWI and non-contrast enhanced MRI is able to replace DCE-MRI, since the acquisition of the DCE-MRI time-series is more time-consuming than the acquisition time of MRI and DWI. Getting accurate diagnostic values in a faster way is always of interest.

A summary of all results is given at the end of this section in Table 6.9.

DCE-MRI feature sets The results are summarized in Table 6.5. The best results are achieved using f_{dce}, followed by $f_{\Delta dce}$. f_{kin} as well as the Haralick texture features are performing worst in average. The boxplot in Figure 6.9 also summarizes the DSC statistics. One box is representing the DSC of all patients for one selected feature subset. The boxplot representation shows that there is a high variability in the DSC among the patients, ranging from 0 up to 0.9. It also shows that no feature subset is able to segment all lesions. There is always at least one lesion with a DSC of 0, indicating that the segmentation failed completely in that case.

In Figure 6.10 the DSC is listed for each patient. The features $f_{\Delta dce}$ seem to improve the result of segmentation in some cases significantly (Pat. 1, 3 and 10) compared to f_{dce}. In other cases the Dice coefficient drops significantly (Pat. 6 and 15). For patient 6 the DSC is going down to 0 for the derivative features. As seen in the first segmentation experiment (Section 6.2.2), the RF mainly picks the derivative features from the early post-contrast phase, where the contrast uptake starts (frames 6-8). Analyzing the time-intensity curve for patient 6 revealed that the contrast uptake starts two frames later than for other patients. The uptake begins at frame 8, for other patients it starts at frame 6. So the RF prediction misses the uptake for this patient, and the segmentation fails using $f_{\Delta dce}$ for this patient.

In Table 6.6 the performance of combinations of DCE-MRI feature subsets is summarized. In Figure 6.11 the DSC is listed for each patient. In average the DCE-MRI feature is performing better than combinations of this feature subset with others.
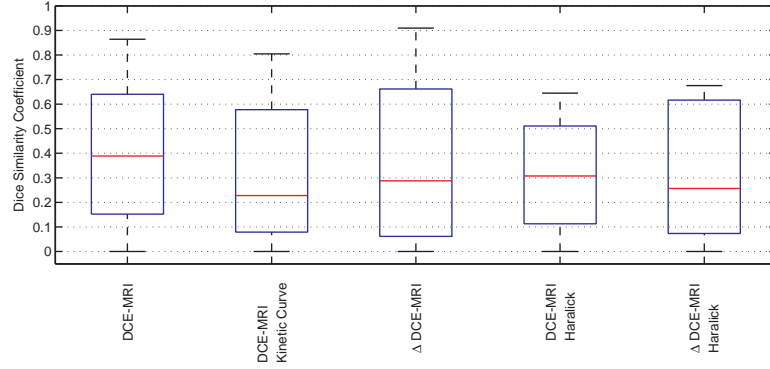
Figure 6.9: DSC boxplot statistics for DCE-MRI feature subsets.
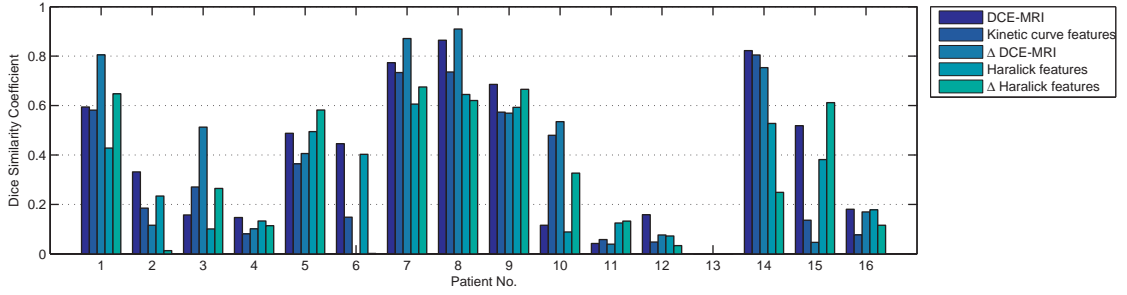
Figure 6.10: DSC listed for each patient. Lesions 1 to 10 are malignant, 11 to 16 are benign lesions.

Feature Type	Mean DSC	Std. DSC
$f_{dce} + f_{\Delta dce}$	**0.3819**	0.3314
$f_{dce} + f_{tex\text{-}dce}$	0.2664	0.2317
$f_{dce} + f_{\Delta tex\text{-}dce}$	0.3571	0.2763
$f_{\Delta dce} + f_{\Delta tex\text{-}dce}$	0.3703	0.2938
$f_{dce} + f_{\Delta dce} + f_{\Delta tex\text{-}dce}$	0.3477	0.2851

Table 6.6: DSC statistics for combinations of DCE-MRI feature subsets.

Multimodal (DCE-MRI, PET, and DWI) feature sets   Table 6.7 and Figure 6.12 summarizes the results using f_{pet}, f_{dwi} and non-contrast enhanced MRI ($f_{dce\text{-}pre}$). Using f_{pet} alone results in the worst performance. A combination of f_{pet} and f_{dwi} increases the segmentation result significantly. Adding $f_{dce\text{-}pre}$ to the features increases the performance in combination with PET and DWI. It can be seen that treating each feature separately does have less predictive power than using them in combination. This is an indication that the modalities provide useful complementary information, at least from the segmentation point of view. The combination of these modalities performs worse than using f_{dce}. It can be concluded that a combination of MRI, DWI, and PET is not an adequate replacement for DCE-MRI.

78
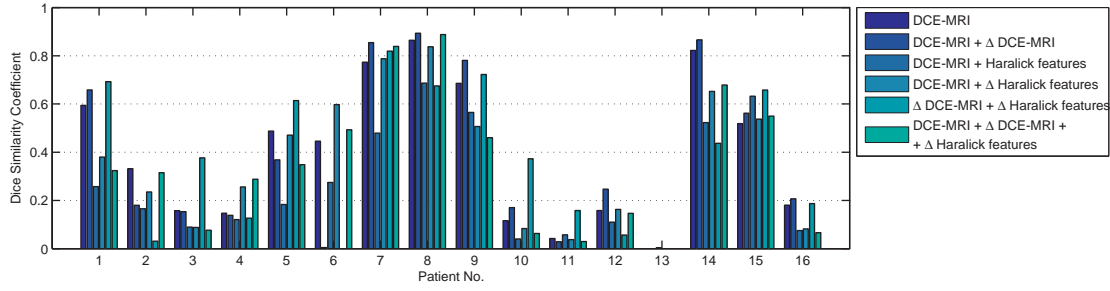
Figure 6.11: DSC of combinations of DCE-MRI feature subsets listed for each patient. Lesions 1 to 10 are malignant, 11 to 16 are benign lesions.

Feature Type	Mean DSC	Std. DSC
f_{dce}	**0.3952**	0.2920
f_{pet}	0.1997	0.2301
f_{dwi}	0.2661	0.2570
$f_{pet} + f_{dwi}$	0.3263	0.2995
$f_{dce\text{-}pre} + f_{pet}$	0.1766	0.2217
$f_{dce\text{-}pre} + f_{dwi}$	0.2587	0.2807
$f_{dce\text{-}pre} + f_{pet} + f_{dwi}$	0.3416	0.3184

Table 6.7: DSC statistics for combinations of PET, DWI and MRI features.
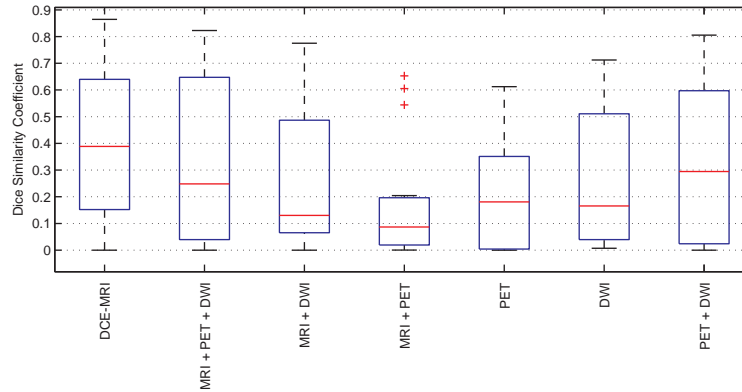
Figure 6.12: DSC boxplot statistics for MRI, PET and DWI feature subsets.

However, the combination of f_{dwi}, f_{pet} and f_{dce} increases the segmentation performance. The results can be found in Table 6.8. The performance of f_{dce} increases from 0.40 to 0.45 when combining it with f_{pet} and f_{dwi}. The performance of $f_{\Delta dce}$ increases from 0.37 to 0.45. Figure 6.13 lists the DSC per patient. For patient 6 it can be observed that f_{dwi} and f_{pet} cannot compensate the segmentation failure of $f_{\Delta dce}$, which has been described above in detail. It can also be noticed that f_{dwi} and f_{pet} improves in common the segmentation performance. But when the segmentation fails completely for DCE-MRI, like in patient 13, adding DWI and PET do not

Feature Type	Mean DSC	Std. DSC
$f_{dce} + f_{pet}$	0.4217	0.2818
$f_{dce} + f_{dwi}$	0.4287	0.3189
$f_{dce} + f_{pet} + f_{dwi}$	0.4486	0.3043
$f_{\Delta dce} + f_{pet}$	0.4562	0.3099
$f_{\Delta dce} + f_{dwi}$	0.4254	0.3234
$f_{\Delta dce} + f_{pet} + f_{dwi}$	0.4540	0.3156

Table 6.8: DSC boxplot statistics for DCE-MRI, PET and DWI feature subsets.
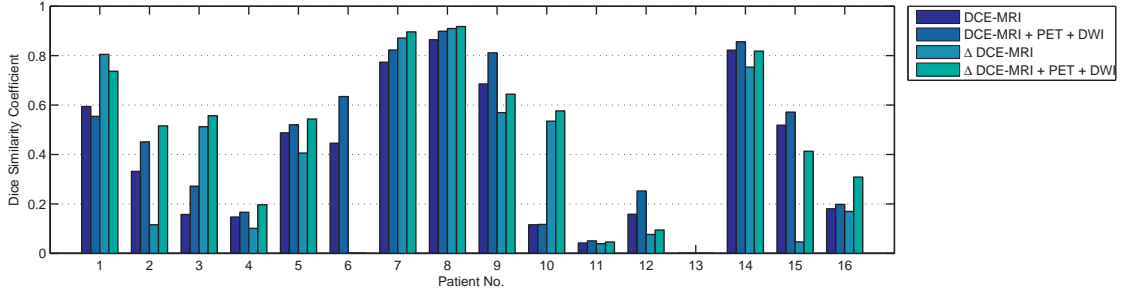
Figure 6.13: Dice similarity coefficient of combinations of DCE-MRI, PET and DWI feature subsets listed for each patient. Lesions 1 to 10 are malignant, 11 to 16 are benign lesions.

improve the segmentation.

Table 6.9 summarizes all results, sorted by the DSC score.

6.2.4 Qualitative Analysis of the Segmentation Results

The previous experiments focused on the quantitative analysis of the segmentation using the DSC. In this section a qualitative analysis is given, where the segmentation results of each lesion are plotted and the advantage and disadvantage of the feature modalities are discussed. In Figure 6.14 the segmentation results for f_{dce} respectively $f_{\Delta dce}$, in combination with f_{dwi} and f_{pet} are plotted for each patient.

It can be seen that using f_{dce} produces more false-positive blobs than $f_{\Delta dce}$ (e.g. patient 1, 3 and 10). On the other hand, when using $f_{\Delta dce}$ it can be observed that the segmentation is often much smaller than the annotated region (patients 2, 4, 6, 12 and 16). For patient 6 even the whole lesion is missed, but this happened because the contrast uptake starts two frames later than for other patients. $f_{dwi} + f_{pet}$ as feature subset produces the noisiest segmentation (e.g. patient 4, 10) and small lesions are completely missed (patients 9, 10, 11, 15 and 16). Since the scanner software does not always align PET and CT image perfectly, an offset in the segmentation can be observed for patients 8 and 14.

Using f_{dce} in combination with f_{dwi} and f_{pet} does not reduce the amount of false positive blobs. But regions inside the lesion, which are missed by f_{dce}, are covered by these combined modalities. This effect is even better visible when combining $f_{\Delta dce}$ with f_{pet} and f_{dwi} (patients 1, 2, 4, 5, 6, 7, 8 and 14). But when f_{dce} fails to segment lesions, combinations with f_{dwi} and

Feature Type	Mean DSC	Std. DSC
$f_{\Delta dce} + f_{pet}$	0.4562	0.3099
$f_{\Delta dce} + f_{pet} + f_{dwi}$	0.4540	0.3156
$f_{dce} + f_{pet} + f_{dwi}$	0.4486	0.3043
$f_{dce} + f_{dwi}$	0.4287	0.3189
$f_{\Delta dce} + f_{dwi}$	0.4254	0.3234
$f_{dce} + f_{pet}$	0.4217	0.2818
f_{dce}	0.3952	0.2920
$f_{\Delta dce} + f_{kin}$	0.3873	0.3536
$f_{dce} + f_{kin}$	0.3818	0.2800
$f_{\Delta dce} + f_{\Delta tex\text{-}dce}$	0.3703	0.2938
$f_{\Delta dce}$	0.3694	0.3385
Automatic feature selection: mRMR 32 bins	0.3624	0.3195
$f_{dce} + f_{\Delta tex\text{-}dce}$	0.3571	0.2763
Automatic feature selection: Random Forest	0.3545	0.3129
$f_{dce} + f_{\Delta dce} + f_{\Delta tex\text{-}dce}$	0.3477	0.2851
$f_{dce\text{-}pre} + f_{pet} + f_{dwi}$	0.3416	0.3184
f_{kin}	0.3297	0.2809
$f_{pet} + f_{dwi}$	0.3263	0.2995
$f_{\Delta tex\text{-}dce}$	0.3158	0.2715
$f_{tex\text{-}dce}$	0.3131	0.2195
$f_{dce} + f_{tex\text{-}dce}$	0.2664	0.2317
f_{dwi}	0.2661	0.2570
$f_{dce\text{-}pre} + f_{dwi}$	0.2587	0.2807
f_{pet}	0.1997	0.2301
$f_{dce\text{-}pre} + f_{pet}$	0.1766	0.2217

Table 6.9: Summary of the DSC statistics, sorted by mean DSC.

f_{pet} are also not able to do the segmentation (patient 13).

It can also be observed that there is often a small offset between the manually annotated border of the lesion and the segmentation (e.g. patients 1, 8 and 14). Especially at the borders of the lesion partial volume effects take place, and it is often not easy to determine for the radiologist, where the exact border of lesion is, resulting in inaccuracies at the lesion border.

Pat No.	f_{dce}	$f_{dce} + f_{pet} + f_{dwi}$	$f_{\Delta dce}$	$f_{\Delta dce} + f_{pet} + f_{dwi}$	$f_{pet} + f_{dwi}$
1					
2					
3					
4					
5					

Pat No.	f_{dce}	$f_{dce} + f_{pet} + f_{dwi}$	$f_{\Delta dce}$	$f_{\Delta dce} + f_{pet} + f_{dwi}$	$f_{pet} + f_{dwi}$
6					
7					
8					
9					
10					
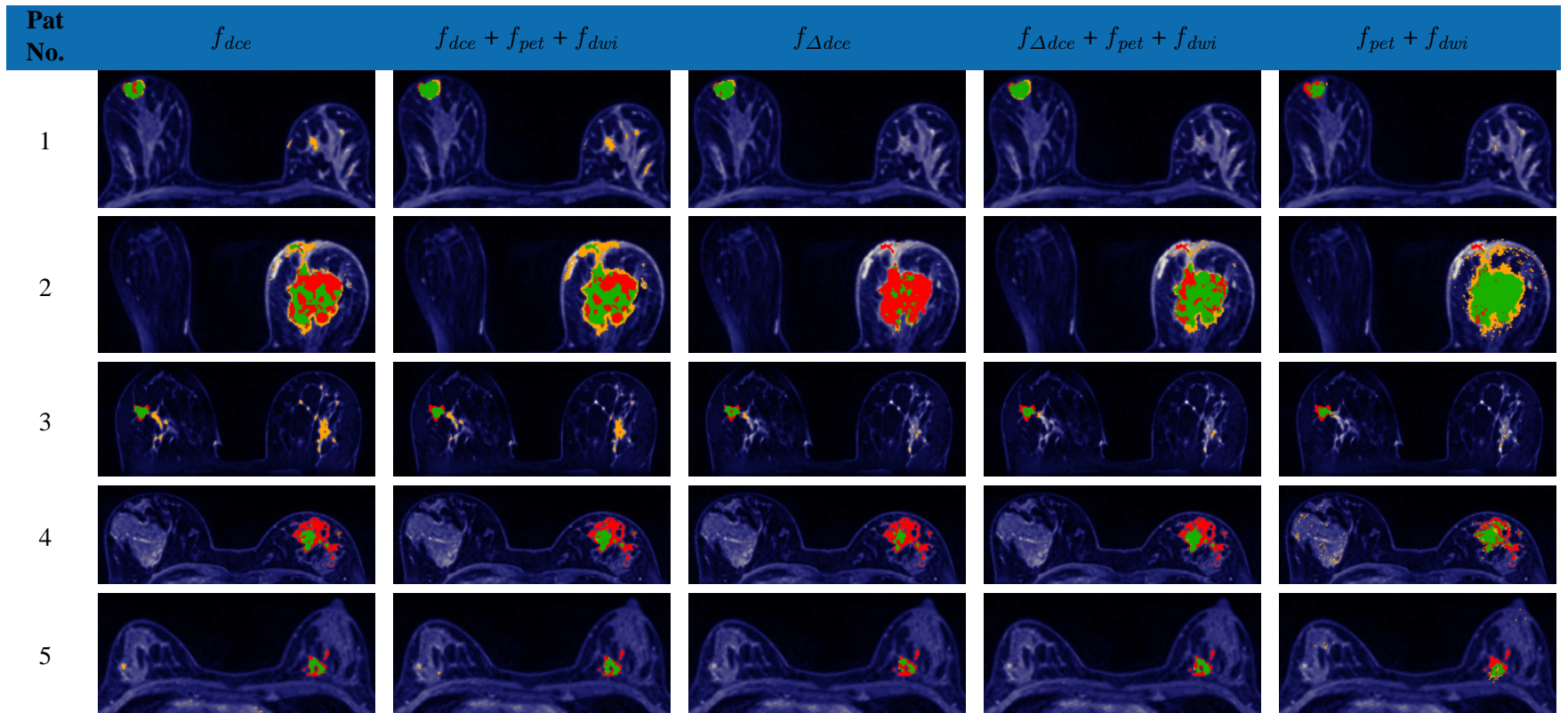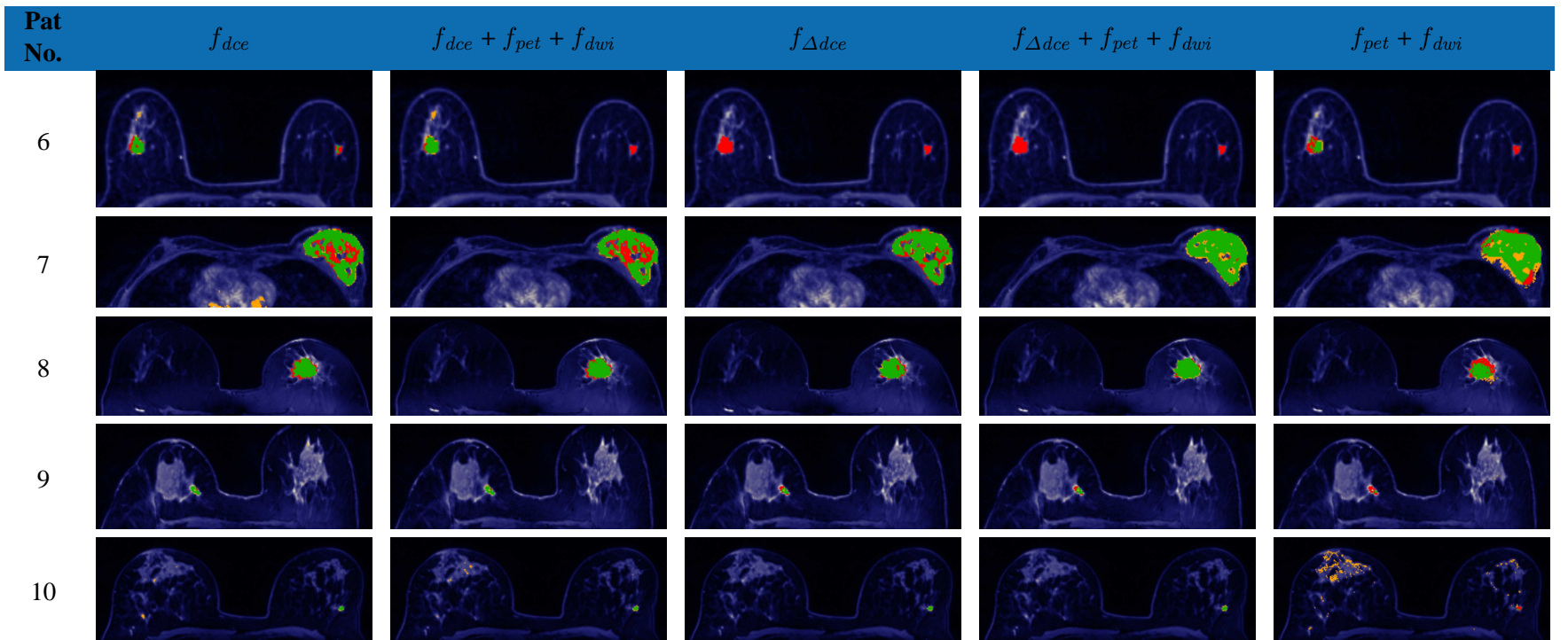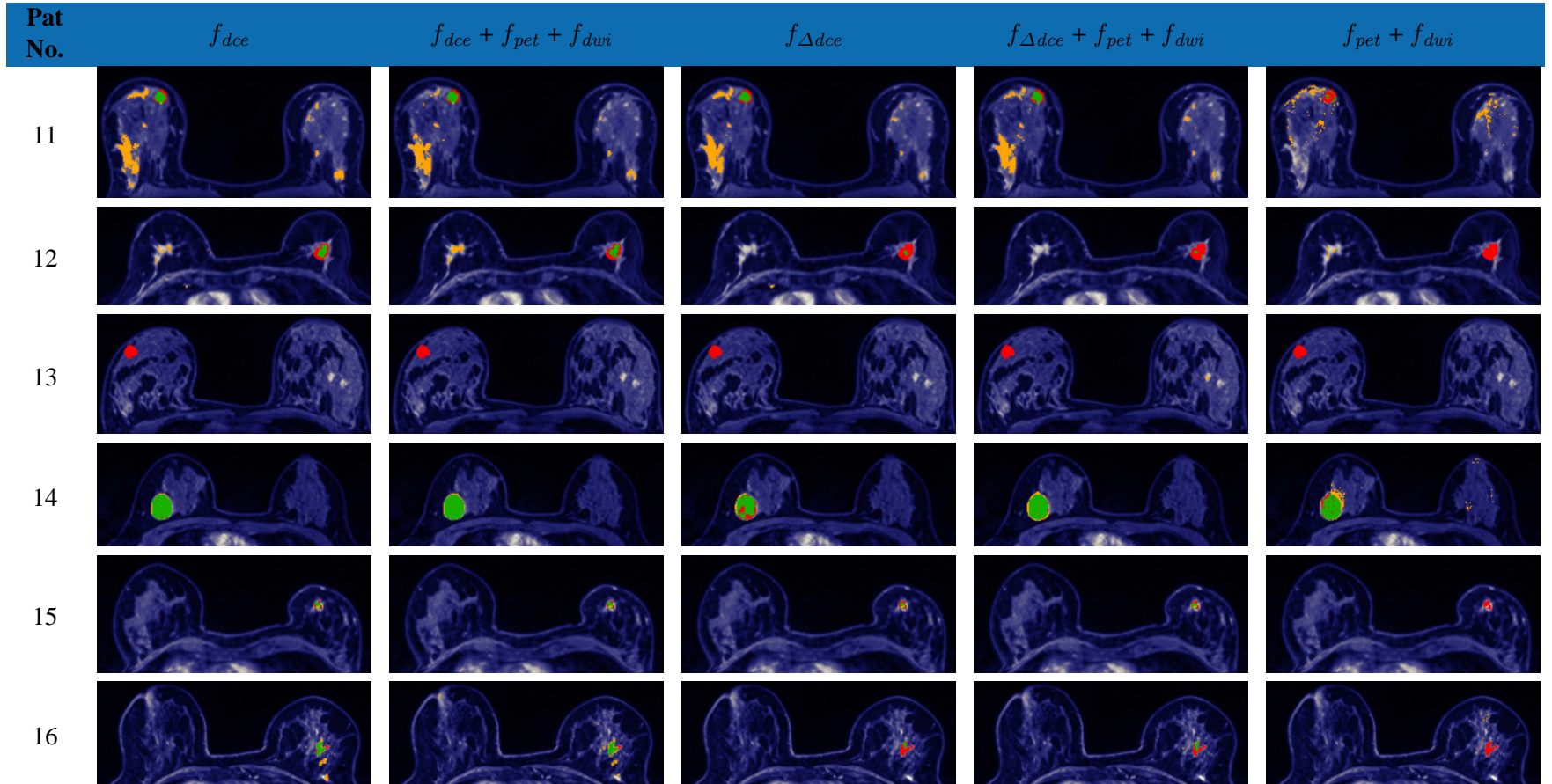
Figure 6.14: Segmentation results for DCE-MRI, PET and DWI features. True positive labels are green, false positive labels are yellow, and false negative labels are of red color. Patients 1 to 10 have malign lesions, patients 11-16 have benign lesions. The order of the patients is the same as in the previous diagrams.

6.2.5 Discussion

In this section the segmentation performance for the modalities MRI, DWI, and PET, as well as the categories DCE-MRI intensities, texture and kinetic curve features has been evaluated. The performance measure has been obtained by calculating the mean DSC score from the manual annotation and the segmentation yielded from the RF training and prediction in a LOOCV process. Furthermore, the Gini importance measure obtained from the RF training has been used to explore the contribution of the features on the segmentation. The ranking obtained from Gini importance was also used for feature selection.

The RF feature selection has been compared with another feature selection method, the mRMR feature selection. These feature selection methods revealed that the segmentation performance remains constant up to a reduction of the feature vector to the top-6-ranked features. The performance of RF and mRMR feature selection was similar, even when a different subset of features was selected. This is an indication that several sets of features share the same information necessary for the segmentation. The top-ranking of DWI in both RF and mRMR feature selection, as well as the significant drop in the performance when removing this feature indicates that f_{dwi} contains valuable information for the segmentation process.

Analyzing the Gini importance features of the DCE-MRI time-series, the secondary texture features, and their derivatives revealed that the relevant information of f_{dce} is in the late post-contrast phase and for $f_{\Delta dce}$ and $f_{\Delta tex\text{-}dce}$ the relevant part is in the early post-contrast phase, where the contrast uptake begins. Although, especially benign lesions can have a slow uptake slope and may be missed by the derivative features, as seen in the qualitative analysis of the segmentations.

By calculating the mean DSC in a LOOCV process following insight has been received: Combining DCE-MRI with the modalities PET and DWI gives the highest score. In general f_{pet} and f_{dwi} improve the segmentation performance when added to f_{dce}. In lesion areas, where DCE-MRI does not enhance (e.g. necrotic areas), and the DCE-MRI segmentation fails for this region, PET and DWI gives the relevant information for correct prediction of these regions. Nevertheless, PET and DWI are not an appropriate replacement for the DCE-MRI modality. Small lesions are often missed by these modalities, due to the relative low resolution of PET and DWI. Also, DCE-MRI in combination with f_{pet} and f_{dwi} fails, when the prediction of the segmentation fails for f_{dce}.

The secondary DCE-MRI features, namely $f_{tex\text{-}dce}$ and f_{kin}, do not improve the segmentation performance in comparison to f_{dce}. Also combinations of them do not improve the performance. This indicates that there is not enough additional information in the secondary features and that they are mostly redundant.

As seen in the qualitative analysis, on the one hand f_{dce} is producing more false positives regions than $f_{\Delta dce}$. On the other hand, f_{dce} also has more true positive regions, so that in sum both feature sets f_{dce} and $f_{\Delta dce}$ have a similar DSC score.

As a side-effect, the experiment revealed that the time-point of the contrast uptake needs to be determined and synchronized for all patient images in following studies.

The segmentations obtained in the previous evaluation are the foundation of the following classification experiments.

6.3 Classification

This part of the evaluation process covers the classification part of the CAD pipeline. The outline of this section is as follows:

- **DWI / PET features**: In the first part several methods of extracting discriminative features from I_{adc} and I_{pet} are presented and evaluated. In particular the discriminative power of FCM clustered I_{adc} values and I_{pet} values is examined, since using FCM to extract relevant I_{adc} values for classification is a novel approach. Furthermore, the FCM clustering is compared with other algorithms extracting I_{adc} values from the segmented lesion.

- **DCE-MRI kinetic curve features**: In the second part the discriminative power of the DCE-MRI kinetic curve features is investigated. Boxplots and the Gini importance give an insight into the relevance of each feature.

- **Classification**: Finally, the classification performance is evaluated in a LOOCV process in two ways. First, the *manual* annotated segmentations of the lesions (L^i) are used, to exclude flaws introduced by the segmentation process and analyze the features, assuming an optimal segmentation. In a second step the evaluation process is performed on the *automatic* segmented lesions (l^i) to investigate how well the full CAD pipeline is performing. The same patient datasets are used for these experiments than in the previous segmentation evaluation.

One may be referred to Table 5.6 in the previous chapter to get an overview of the classification features used in the following experiments.

Experimental Setup   The subsequent experiments share following experimental setup where not otherwise stated. Data from the same 16 patients as in the previous segmentation experiments are used. The lesions of 10 patients are malignant and 6 are benign. The histopathologically report of the lesions is considered as the ground truth.

The classification features are calculated from the voxels within the lesion. The lesion region is specified either by the manually annotated lesions L^i, or by the segmentation l^i obtained from the segmentation step in the CAD pipeline. For the automatic segmentation the feature combination f_{dce}, f_{pet} and f_{dwi} has been used. This feature subset got the third-highest ranking in the segmentation evaluation process, after the $f_{\Delta dce}$, f_{pet}, f_{dwi} variants. These choice of features produces on the one hand more false positive blobs than the $f_{\Delta dce}$ feature subsets, on the other hand f_{dce} covers the lesion better than their derivatives. In a post processing step false positive blobs are removed. The segmentation is divided into blobs using an 8-neighborhood. The blobs covering L^i are used for classification process, and the other blobs are dropped. This step simulates a manual selection of a suspicious region or blob by a radiologist, which he wants to investigate further. One lesion, which was not detected in the segmentation process, has been added to the classification evaluation, due to the low amount of available patients for the classification evaluation. For this patient the manual annotated segmentation is used in both cases. This patient is listed as number 13 in the segmentation experiments above.

The *mTry* parameter, defining the number of feature variables tried at each split in the RF algorithm is set to $\lfloor \sqrt{\# \text{ features}} \rfloor$ as suggested by Breiman [28].

6.3.1 Evaluation of the DWI-ADC and PET features obtained via Fuzzy C-means Clustering Method

The aim of this test is to examine, if the FCM clustering of I_{adc} and I_{pet} value gives discriminative features for the classification of the lesion. The FCM algorithm is compared with other methods of determining the "best" DWI-ADC values within a lesion.

The algorithms should imitate up to a certain degree the ROI selection by a radiologist in a manual or semiautomatic classification process. In such cases the radiologist draws the ROI inside the lesion, choosing a region with a low DWI-ADC value. Studies demonstrated that this procedure gives good discriminative features [24]. In a fully automatic classification process the ROI has to be defined by an algorithm. The algorithm should select within the lesion a region with a discriminative DWI-ADC value and it should be robust against outliers introduced by inaccuracies at the lesion border, noise in the data, and flaws in the segmentation process.

Following algorithms have been chosen to get the DWI-ADC value:

- the minimum value within the lesion,

- the minimum value of 3x3x3 median filtered ADC values within the lesion,

- a FCM clustering with fixed cluster size,

- a FCM clustering with variable cluster size and

- a k-means clustering with fixed cluster size.

The PET values are evaluated using the same algorithms, with the only difference that the maximum value is used instead of the minimum value.

Experimental Setup   The experiment is performed on the manual annotated segmentations of all 16 patients, to eliminate possible errors introduced by an inaccurate segmentation process. The ADC value is calculated for each patient using each one of the five algorithms.

The number of clusters is a crucial parameter in k-means and FCM clustering. Therefore, the k-means clustering and the FCM clustering are performed with number of clusters ranging from 2 to 15. The cluster center with the lowest ADC value is chosen as the representative ADC value of the lesion.

In addition, a FCM clustering with a variable number of clusters has been evaluated, where the number of clusters is dependent on the size of the lesion. This method is inspired by the FCM clustering of the DCE-MRI intensity curves to get the CKC, as proposed by Chen et al. [36]. The number of clusters c is in this case dependent on the number of voxels N: $c = max([N/80], 2)$.

To compare the automatic approach with a manual approach, a radiologist drew a squared ROI within the lesion, from which an average ADC value has been calculated. The ROIs were drawn in the unregistered DWI ADC images.

Validation Measures The statistics of the ADC values from the 16 patients are summarized in a boxplot grouped by each algorithm. So the mean, variance, and other statistical properties and the discriminative power of each algorithm can be visually examined.

Results In Figure 6.15 the summary of the ADC values are plotted using boxplots. Choosing the minimum of the ADC values is very sensitive to outliers, and often an ADC value of 0 is chosen. The minimum of the median filtered values is less sensitive to small outliers. The ADC values do not perfectly discriminate the malign from the benign lesions; there is an overlap in the values. A nearly perfect discriminative result is given by using FCM clustering with 4 clusters. There is one outlier in the benign ADC values, which is in the region of the malign ADC values. The discriminative power of FCM clustering with variable cluster size is worse than using a fixed cluster size. Using a k-means clustering with 4 clusters gives comparable results than the FCM clustering, with a slightly higher variance of the clustered ADC data values. The ADC value from the manually selected ROI are perfectly discriminating the lesion classes.
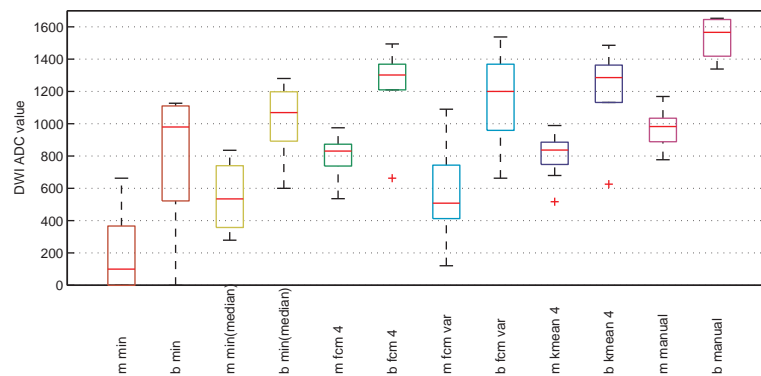
Figure 6.15: ADC values of the 16 patients summarized by boxplots and grouped by malign (m) and benign (b). The algorithms used for calculation are from left to right: minimum value, minimum of median filtered values, FCM clustering with 4 clusters, FCM clustering with variable cluster size, k-means clustering and data from manual drawn ROI

Figure 6.16 shows a boxplot summary of the ADC values for a cluster size ranging from 2 to 15. The ADC values are already discriminative when using 2 clusters. The variance of the clustered ADC values is increasing with a higher cluster size. The maximum gap between the benign and malign ADC values, ignoring the outlier, is achieved when using cluster sizes between 3 and 5.

The FCM clustering for one lesion is visualized in Figure 6.17. The histogram reveals that there are a few outliers with an ADC value of 400. These ones are chosen when using the minimum of ADC values. The main amount of ADC values lies within 700 to 1400, which is mainly covered by the first two of the 4 clusters.

The algorithms are also applied on I_{pet}, using the maximum of the intensity values. Figure 6.18 presents the results of the achieved PET values. The majority of the malign lesions have a higher PET intensity value than the benign lesions, whereas the discriminative power is not
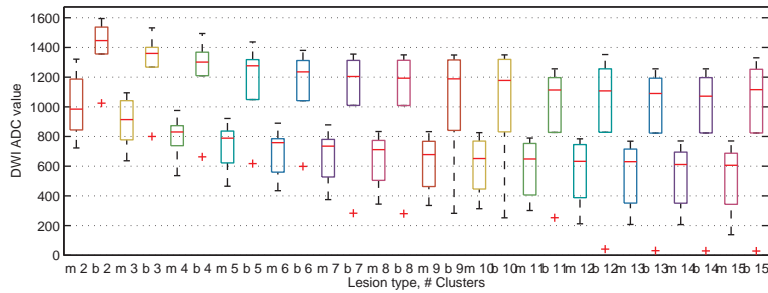
Figure 6.16: Boxplot summary of ADC values for given cluster size, grouped by malign (m) and benign (b) lesions.
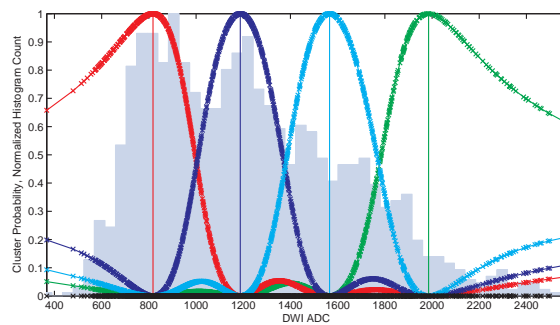
Figure 6.17: FCM clustering of ADC values for one lesion. The colored curves are the probabilities, that a ADC value belongs to a cluster. The vertical lines represent the cluster centers. The histogram in the background shows the distribution of the voxel ADC values.

as good as for ADC values due to overlaps of the data. The PET clustering algorithm does not contrast from the other algorithms as much as it did for ADC-DWI clustering.

Figure 6.19 visualizes the centroid PET values for a cluster size ranging from 2 to 15. The clustered PET value rises with increasing number of clusters, but the values are already discriminative when using 2 clusters.

6.3.2 Evaluation of the DCE-MRI Kinetic Curve Features

The kinetic curve features extracted from the fitted asymmetric generalized logistic function are examined in this experiment, which are MaxEnh, TTP, MDER, and AuC. For each of these features the discriminative power is analyzed using boxplots and the Gini importance from a trained RF.

Experimental Setup To get the importance ranking of the features a RF training is started using all curve features obtained from the manual annotation, respectively the automatic segmentation. The tree size $nTree$ has been set to 1000.
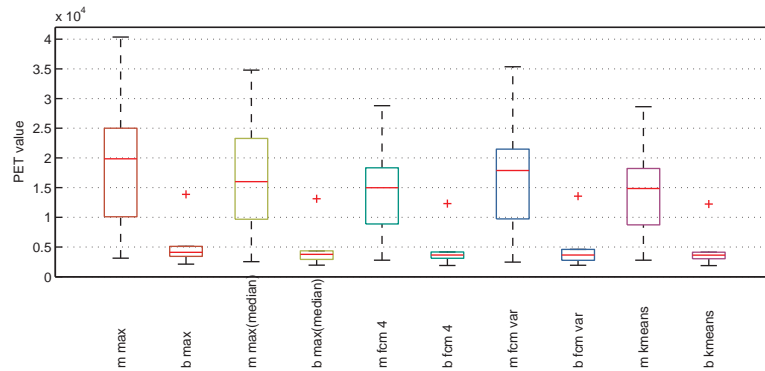
**Figure 6.18:** PET intensity values of the 16 patients summarized by boxplots and grouped by malign (m) and benign (b). The algorithms used for calculation are from left to right: maximum value, maximum of median filtered values, FCM clustering with 4 clusters, FCM clustering with variable cluster size, k-means clustering with 4 clusters.
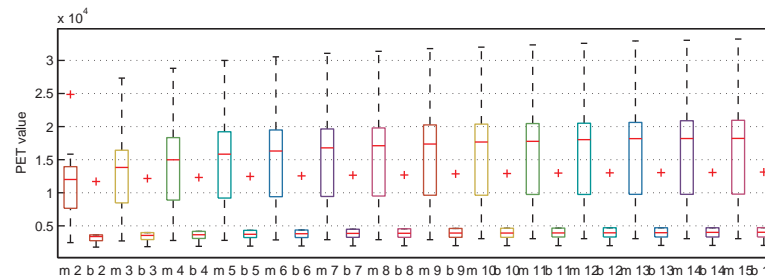
**Figure 6.19:** Boxplot summary of PET values for given cluster size, grouped by malign (m) and benign (b) lesions.

Validation Measures   The discriminative power of the features is visualized by using boxplots. For each feature the patient data is divided into benign and malignant groups and a boxplot statistic is calculated from the feature values. The importance of a feature is indicated by the Gini importance, yielded from the RF training.

Results   The distribution of the feature data is summarized in Figure 6.20 by a boxplot diagram. The distribution is plotted for both data pools, the one from the manual annotation and the one from the automatic segmentation. The boxplots visualizes how well the features are discriminative when treating them independently from each other. The MaxEnh and the AuC is in the same range for benign and malign lesions, with a higher variance in the benign data. The TTP feature reveals that the peak of the curves for the benign lesions is always at the end of the DCE-MRI sequence, which indicates a constant increase of the signal intensity. For malignant lesions the TTP values ranges from the early postcontrast phase to the late postcontrast phase. The mean of the MDER feature is higher for malign lesions, which implies a rapid contrast
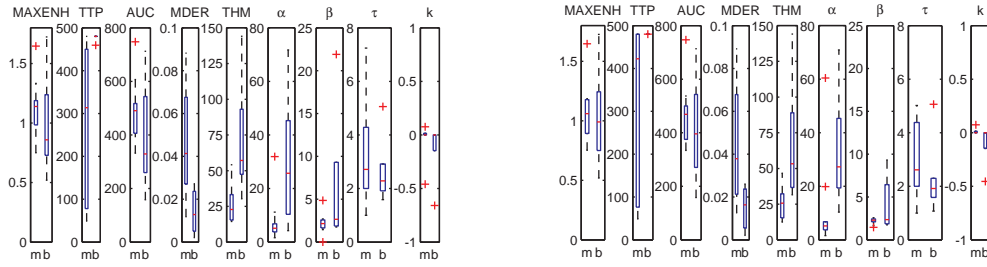
Figure 6.20: Boxplots of the curve features used for classification, grouped by lesion type malignant (m) and benign (b). Left: Data taken from manual annotated lesions, right: Data from segmented lesion.

enhancement in the malign lesions. The mean THM is higher for benign lesions, which is an indication that the contrast uptake of benign lesions is slower than for malign lesions. These observations are in concord with the results of Kuhl et al. [85, 87], where it has been observed that 83% of the benign lesions have a constant increase of the signal intensity, and that a washout in the late post-contrast phase in combination with a rapid and intense enhancement is indicative for malignancy.

In Figure 6.21 the Gini importance resulting from the RF training is plotted. For the manual
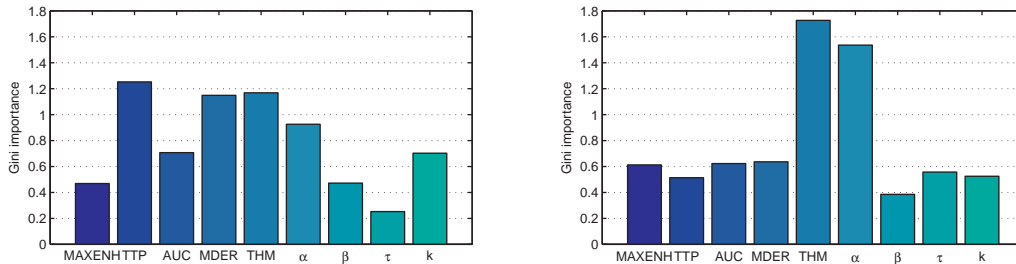
Figure 6.21: The Gini importance of the DCE-MRI kinetic curve features. Left: Data taken from manual annotated lesions, right: Data from segmented lesion.

annotated lesions the TTP, MDER and THM features get the highest ranks. For the automatically segmented lesions the α and THM feature dominates the others features. Examining the boxplots from Figure 6.20 reveals that there is less overlap of the benign and malign α values obtained from the automatically segmented lesion than for the α values obtained from the manual annotations, which results in a higher discriminative power and finally in a higher importance score of the RF. MaxEnh and AUC, as well as β and τ have a low score in both cases. As seen in the boxplots there is a large amount of overlap for these data values, and this makes it less attractive for the RF, resulting in a lower score.

6.3.3 Classification Performance for Feature Subsets

The following experiment is evaluating the classification performance using RF as a classifier. The benefits of each modality are evaluated by trainining and prediction of the class using subset of features and combinations of them. The features are divided into following subsets:

- $f_{l\text{-}adc}$, $f_{l\text{-}pet}$: The DWI-ADC and PET value from FCM clustering with 4 clusters.

- $f_{l\text{-}curve}$: features from the fitted asymmetric logistic function (α,β,τ, k, AUC, THM, TTP, MDER, MAXENH).

- $f_{l\text{-}ckc}$, $f_{l\text{-}\Delta ckc}$: the plain DCE-MRI CKC intensity values, as well as their derivatives.

Experimental Setup   The classification of the lesion is done using a RF classifier. The number of trees (*ntree*) in the RF training is set to 200.

Validation Measures   The performance is evaluated using a LOOCV process with 20 iterations. As performance measure the sensitivity, specificity, correct rate, error rate, positive predictive value, negative predictive value, AuC, and the F-measure is calculated. A short explanation of these performance measures is given in the Appendix Table A.2. The tests are performed using the data obtained from the manual annotation L^i, as well as the data obtained from the automatic segmentation l^i.

As rule for the RF votes on the one hand the majority rule is used, where more than half of the votes of a RF prediction must be assigned to one class, to decide for this class. On the other hand an optimal threshold value for the votes is calculated by maximizing the F-measure, which is the harmonic mean of sensitivity and specificity. The threshold value is iterated from 0 to 200 (ntree) and the F-measure is calculated for the given threshold value. The results given by the threshold value with the maximum F-measure, are stored.

The Receiver Operating Characteristic (ROC) curve [112], on which the AuC measure relies on, is computed by varying the threshold level for the votes from 0 to *ntree* and calculating the sensitivity and (1-specificity) using the given threshold.

6.3.4 Results

Table 6.10 summarizes the F-Measure, the sensitivity, the specificity and the AuC of the ROC curve for feature combinations using the majority rule and an optimized threshold value. A complete summary with all of the performance measures listed above, is given in the Appendix Tables A.3 and A.4.

Choosing an optimal threshold increases the prediction performance significantly in all cases. For instance, the F-Measure for $f_{l\text{-}adc}$ increases from 0.82 to 0.90 and the sensitivity/specificity from 0.80/0.83 to 1.0/0.83 when using an optimal threshold. The trade-off between sensitivity and specificity can be observed in these two tables. When using an optimal threshold the sensitivity is in common decreasing and the specificity is increasing.

$f_{l\text{-}adc}$ is performing best in all cases when comparing single feature subsets only ($f_{l\text{-}adc}$, $f_{l\text{-}ckc}$, $f_{l\text{-}\Delta ckc}$, $f_{l\text{-}curve}$, and $f_{l\text{-}pet}$). The sensitivity of 1.0 and specificity of 0.83, using an optimal

threshold reflects the distribution of the feature dataset as observed in Figure 6.15. The clustered ADC data values are perfect distinct, with the exception of one outlier. The performance is similar for both manual annotated lesion data and automatic segmented data. $f_{l\text{-}adc}$ and its combination with other features are always in the top-ranks, both for manual annotation and automatic segmentation as well as for majority votes and optimal threshold value. Using $f_{l\text{-}ckc}$, $f_{l\text{-}pet}$ or $f_{l\text{-}curve}$ features without $f_{l\text{-}adc}$ for classification results in a low F-Measure and a low ranking in the table.

There is a performance drop observed between manual and automatic segmented data, when using $f_{l\text{-}ckc}$ or $f_{l\text{-}\Delta ckc}$, also in combination with other features. On the other hand, the F-value and AuC measure for $f_{l\text{-}curve}$ in combination with other feature sets are higher for the segmented data when using an optimal threshold.

Focusing on the AuC, which is independent of the choice of the threshold, it can be observed that the measure increases when modalities are combined. For instance, for automatic segmentation the AuC for $f_{l\text{-}adc}$ is 0.80, for $f_{l\text{-}pet}$ 0.59, and for $f_{l\text{-}curve}$ 0.65. Though, when combining $f_{l\text{-}adc}$ with $f_{l\text{-}pet}$ and $f_{l\text{-}curve}$, a top score of 0.94 is obtained for the automatic segmentation, and $f_{l\text{-}adc} + f_{l\text{-}curve}$ results in the top AuC score of 0.90 for manual segmentation data. This indicates that the robustness of the classifier increases when using combinations of several modalities.

6.3.5 Discussion

In this section the classification performance of the proposed CAD system has been evaluated, as well as an insight in the proposed classification features have been given. In the first part the proposed FCM clustering algorithm has been examined, where I_{adc}, respectively I_{pet} values within the lesion are clustered and the cluster center with the lowest, respectively highest value is picked as representative value. FCM has been compared with the algorithms of minimum value selection of raw data and mean filtered data, as well as k-means clustering. The experiments revealed that FCM clustering of the ADC values with a cluster size greater than 2 gives discriminative features. In the second part of this section the kinetic curve features extracted from the regression curve have been examined, revealing on the one hand that there is a variance in the values obtained from the manual annotated data pool and the automatic segmented data, resulting in a higher performance in the final classification for the automatic segmentation.

Finally, the performance of the classifier using subsets of features has been measured. The results showed that when using a combined feature set of all three modalities, the performance surpasses the performance of a single modality approach. The AuC values for $f_{l\text{-}adc}$, $f_{l\text{-}pet}$, and $f_{l\text{-}ckc}$, respectively $f_{l\text{-}curve}$ are for the automatic segmented lesions: 0.80, 0.59, 0.51, respectively 0.65, whereas the combination of $f_{l\text{-}adc}$, $f_{l\text{-}pet}$, and $f_{l\text{-}curve}$ gives an AuC of 0.94. This indicates that the modalities provide complementary information for the classification, resulting in a higher score.

Although, due to the small sample size of 16 patients the expressive power of the statistics is quite limited. The increase in performance may therefore be seen more as a trend than as a solid statistical performance measure. Anyway, with an increase of the sample size similar results are expected, since the results are compliant with results from other studies. Yabuuchi et al. [172] reports a sensitivity and specificity of 82% and 86% for combined DWI-ADC, DCE-MRI morphological and kinetic curve features setup, where the data is interpreted by radiologists. In this

thesis the sensitivity and specificity for combined DWI-ADC and DCE-MRI kinetic curve features is 87%/99% for the automatic segmentation, respectively 81%/91% for manual annotated lesions. Levman et al. [91] used DCE-MRI kinetic curve features in a CAD system and reports an AuC, respectively sensitivity and specificity of 0.74, 63% and 79%. In this study the DCE-MRI kinetic curve features shows an AuC, sensitivity and specificity of 0.70, 67% and 83% for the manual annotated lesions, respectively 0.65, 61% and 68% for the automatic segmentation.

The proposed CAD system gives comparable results to state-of-the-art DCE-MRI CAD systems. In a recent study Agner et al. [2] list an AuC, sensitivity and specificity of 0.92, 95% and 82% using morphological and spatiotemporal texture features. One may be referred to Table 4.1 in the "Recent Work" chapter to get the performances of the various DCE-MRI CAD studies.

Features	F-Measure		Sensitivity/Specificity			
	man	seg	man		seg	
$f_{l\text{-}adc} + f_{l\text{-}pet}$	**0.83**	0.75	0.90	0.77	0.86	0.68
$f_{l\text{-}adc}$	0.82	**0.82**	0.80	**0.83**	0.80	**0.83**
$f_{l\text{-}adc} + f_{l\text{-}\Delta ckc}$	0.73	0.28	0.82	0.66	0.80	0.17
$f_{l\text{-}adc} + f_{l\text{-}curve}$	0.70	0.59	0.90	0.57	**0.90**	0.44
$f_{l\text{-}adc} + f_{l\text{-}ckc} + f_{l\text{-}\Delta ckc}$	0.69	0.31	0.92	0.55	0.83	0.19
$f_{l\text{-}adc} + f_{l\text{-}pet} + f_{l\text{-}curve}$	0.68	0.66	0.90	0.55	**0.90**	0.53
$f_{l\text{-}\Delta ckc}$	0.66	0.02	0.81	0.56	0.75	0.01
$f_{l\text{-}ckc} + f_{l\text{-}\Delta ckc}$	0.66	0.33	**0.94**	0.51	0.70	0.22
$f_{l\text{-}adc} + f_{l\text{-}ckc}$	0.63	0.36	0.80	0.52	0.80	0.23
$f_{l\text{-}pet} + f_{l\text{-}curve}$	0.63	0.61	0.84	0.50	**0.90**	0.47
$f_{l\text{-}ckc}$	0.62	0.46	0.80	0.50	0.72	0.33
$f_{l\text{-}pet} + f_{l\text{-}ckc}$	0.60	0.47	0.75	0.50	0.79	0.33
$f_{l\text{-}curve}$	0.57	0.49	0.80	0.44	**0.90**	0.33
$f_{l\text{-}pet}$	0.2692	0.27	0.70	0.17	0.70	0.17

Features	F-Measure		Sensitivity/Specificity				AuC	
	man	seg	man		seg		man	seg
$f_{l\text{-}adc}$	**0.91**	0.91	**1.00**	0.83	**1.00**	0.83	0.79	0.80
$f_{l\text{-}adc} + f_{l\text{-}\Delta ckc}$	0.89	0.73	0.80	**1.00**	0.66	0.82	0.87	0.66
$f_{l\text{-}adc} + f_{l\text{-}pet}$	0.87	0.82	0.92	0.83	0.82	0.83	**0.90**	0.89
$f_{l\text{-}adc} + f_{l\text{-}curve}$	0.86	0.93	0.81	0.91	0.87	**0.99**	0.88	0.91
$f_{l\text{-}\Delta ckc}$	0.84	0.60	0.76	0.95	0.58	0.63	0.79	0.45
$f_{l\text{-}adc} + f_{l\text{-}ckc}$	0.81	0.70	0.79	0.83	0.64	0.77	0.78	0.69
$f_{l\text{-}adc} + f_{l\text{-}pet} + f_{l\text{-}curve}$	0.81	**0.94**	0.83	0.79	0.89	**0.99**	0.86	**0.94**
$f_{l\text{-}adc} + f_{l\text{-}ckc} + f_{l\text{-}\Delta ckc}$	0.80	0.58	0.88	0.73	0.49	0.73	0.81	0.53
$f_{l\text{-}ckc} + f_{l\text{-}\Delta ckc}$	0.76	0.52	0.78	0.75	0.53	0.52	0.75	0.38
$f_{l\text{-}pet}$	0.75	0.76	0.69	0.83	0.70	0.83	0.59	0.59
$f_{l\text{-}ckc}$	0.74	0.55	0.70	0.79	0.62	0.49	0.70	0.51
$f_{l\text{-}curve}$	0.74	0.64	0.67	0.83	0.61	0.68	0.70	0.65
$f_{l\text{-}pet} + f_{l\text{-}ckc}$	0.72	0.54	0.70	0.74	0.75	0.42	0.71	0.55
$f_{l\text{-}pet} + f_{l\text{-}curve}$	0.70	0.76	0.76	0.66	0.84	0.70	0.72	0.77

Table 6.10: Classification performance results using LOOCV with 20 iterations. The upper Table represents the results using a majority vote for classification, and the lower Table the results using an optimized threshold value. Performance values are listed for the manual annotated lesions L^i (*man*) and the automatically segmented lesions l^i (*seg*). The feature subset contains: DWI-ADC value from FCM clustering ($f_{l\text{-}adc}$), PET values from FCM clustering ($f_{l\text{-}pet}$), data values from the Characteristic Kinetic Curve (CKC) and their derivatives ($f_{l\text{-}ckc}$, $f_{l\text{-}\Delta ckc}$), and features from the fitted asymmetric logistic function ($f_{l\text{-}curve}$). The table is sorted by the F-Measure of the manual annotated lesions. The AuC is identically for both majority vote threshold and optimal threshold, and is therefore listed only once.

CHAPTER 7

Conclusion

In this final chapter the crucial points of the Thesis are recapitulated and a summary of the proposed CAD pipeline is given. A few ideas for possible future work are also given.

7.1 Summary

In this thesis a CAD system has been proposed, providing fully-automatic segmentation and classification of breast lesions using a novel multimodal imaging approach. It has been hypothesized that the modalities DCE-MRI, DWI, and PET provide complementary information, resulting in an improvement of the CAD system. A main point of the thesis is the comparison of the multimodal approach with single modality approaches. The performance of the CAD system has been evaluated and compared with several state-of-the-art CAD systems.

The proposed CAD pipeline consisted of three main steps: registration of the modalities, segmentation of the lesions, and the classification of the lesion into being benign or malignant.

In the registration step the image modalities were transformed to a reference coordinate system, so that the spatial position of a voxel is equal for all modalities. Patient movement, distortions in the DWI modality, as well as the differing patient positions in the MRI and PET scanners had to be compensated. A LDDMM registration approach, the Symmetric Normalization [14], in combination with CC and MI similarity metrics was able to register these modalities. The PET modality has been registered indirectly by registering the CT modality acquired in PET-CT scanner to the MRI modality, and applying the yielded transformation on the PET image. Using combined PET-MRI scanners, which are currently in the prototypical phase, will render this indirect registration step obsolete.

In the segmentation step each voxel within the breast has been classified as either cancerous or non-cancerous by a RF. Intensity based, textural based, and kinetic curve features have been proposed. The evaluation of the segmentation revealed that the textural and kinetic curve features did not improve the performance of the segmentation. The evaluation also revealed that the multimodal approach improves the segmentation performance in comparison with a single-modality DCE-MRI approach. 15 from 16 lesions had been successfully located by the

97

algorithm, although the segmentation performance with a mean DSC of 0.45 using DCE-MRI, DWI and PET intensity values needs to be improved in future studies. Ideas of getting a better segmentation are listed in the next section.

The classification step was the final step in the CAD pipeline, where the segmented lesions had been classified into malign and benign using RF. For this purpose the DCE-MRI, DWI, and PET data within the lesion have been clustered using FCM, where clustering DWI and PET is a novel approach. In the evaluation process it has been shown that the clustering process gives discriminative features. It has been also revealed that the multimodal approach surpasses the classification performance using only a single modality.

The main contribution of the thesis was the automatic fusion of multimodal PET, MRI, and DWI breast image data by a registration process, the usage of the fused information from all modalities for automatic segmentation and classification of the lesion, the usage of RF for detecting significant information for segmentation and classification in the multimodal dataset, and the FCM clustering method for automatic selection of significant regions in DWI and PET modalities for lesion classification.

It can be concluded that the proposed CAD pipeline gives comparable results to other state-of-the-art CAD methods. The hypothesis that the modalities provide complementary information, resulting in a performance improvement, has been validated and confirmed in the evaluation process.

7.2 Future Work and Improvements

The approach described in this thesis gives plenty of opportunities for continued research and improvement, which are exemplarily addressed in this section.

Modalities: Initial studies using Magnetic Resonance Spectroscopy for lesion diagnosis showed promising results [63]. This modality can be easily added to the proposed CAD system. Similar evaluation methods as proposed in this Thesis may be applied on the additional modality, in order to analyze its potential in a CAD system.

Registration: The LDDMM registration process is computational demanding, resulting in registration times of more than 3 hours for a single patient. This limits the potential of the CAD system in the routine clinical practice. However, there is room of optimization in the registration framework. The version of the ANTS registration framework used in this Thesis is single-threaded, computing the registration on a single core only, even when there are more available. Adapting the algorithm to be multi-threaded would decrease the computation time significantly [121]. The switch from the CPU to the GPU may become also of interest, which enables massive parallel computing.

During the work on this Thesis it has been observed that the CT/PET scanner software does not always align PET and CT image perfectly. An additional rigid registration using MI may align them better and increase the segmentation performance.

98

Segmentation: The segmentation as proposed in this thesis is performed on a voxel-level, ignoring structural and topological relations. Methods, which consider the spatial relationship, like active contour [3], graph-cut [157], or superpixel segmentation [127] in combination with the multimodal dataset may improve the segmentation performance.

An important point in future studies has to be that the time-point of the contrast uptake is determined and synchronized for all patient images when using the contrast uptake as segmentation feature.

Classification: The classification features proposed in this thesis relies on image intensities and intensity kinetics only. Other current CAD systems include textural and morphological features too [2,100,111,181]. Adding these features may increase the classification performance.

Since producing manual segmentations of breast lesions is a time-consuming and unsatisfying process, a multiple-instance-learning (MIL) approach [46] may be an interesting alternative in the context of breast cancer detection. Instead of using instances with segmented and labeled lesions for training, labeled *bags* containing many instances are used for training. As label the classes 'no lesion' and 'lesion' may be used. From such a collection of labeled bags, the classifier tries to induce a model, which can be used to label the individual instances correctly.

Tables

A.1 Haralick Texture Features

The Haralick texture features are calculated from the GLCM as follows:

Notation N is the number of distinct intensities in the image. $p(i,j)$ is the (i,j)th entry in the normalized GLCM.

Following auxiliary statistic parameters are calculated to simplify the Haralick feature calculation: p_x and p_y are the marginal distributions. They are obtained by summing over the rows, respectively summing over the columns (Eqn. A.1). μ_x and μ_y are the marginal means of the two random variables and are calculated as described in Equation A.2. σ_x and σ_y are the variances of the two random variables (Eqn. A.3). The sum marginal probability p_{x+y} is defined in Equation A.4) and the difference marginal probability p_{x-y} is defined in Equation A.5).

$$p_x(i) = \sum_j p(i,j) \qquad\qquad p_y(j) = \sum_i p(i,j) \qquad\qquad \text{(A.1)}$$

$$\mu_x = \sum_i p_x(i) \qquad\qquad \mu_y = \sum_j p_y(j) \qquad\qquad \text{(A.2)}$$

$$\sigma_x^2 = \sum_i (i - \mu_x)^2 p_x(i) \qquad\qquad \sigma_y^2 = \sum_i (i - \mu_y)^2 p_y(i) \qquad\qquad \text{(A.3)}$$

$$p_{x+y}(n) = \sum_i \sum_j p(i,j) \qquad\qquad \text{with } n = i + j \qquad\qquad \text{(A.4)}$$

$$p_{x-y}(n) = \sum_i \sum_j p(i,j) \qquad\qquad \text{with } n = |i - j| \qquad\qquad \text{(A.5)}$$

F#	Feature	Formula	Description
f_1	Energy (Angular Second Moment)	$$\sum_i \sum_j p(i,j)^2$$	Measures the number of repeated pairs. A high Energy indicates a high occurrence of similar pixel pair intensities.
f_2	Contrast (Inertia)	$$\sum_i \sum_j (i-j)^2 p(i,j)$$	Measures the contrast of the pixel pairs. A large Contrast value indicates a large difference of the intensities.
f_3	Correlation	$$\frac{1}{\sigma_x \sigma_y} \sum_i \sum_j (i-\mu_x)(j-\mu_y)p(i,j)$$	Measures the correlation of two variables by using the covariance values. The Correlation is high if the intensities of the pixels are highly correlated.
f_4	Sum Of Squares: Variance	$$\sum_i (i-\mu_x)^2 p_x(i)$$	Measures the variance of the distribution of the intensity values.
f_5	Inverse Difference Moment (Homogeneity)	$$\sum_i \sum_j \frac{p(i,j)}{1+(i-j)^2}$$	Measures the homogeneity of the pixel pairs. The Homogeneity is large if the intensities of the pairs are similar.

F#	Feature	Formula	Description
f_6	Sum Average	$$\sum_{n=2}^{2N} n p_{x+y}(n)$$	Measures the mean of the intensity values. A high Sum Average corresponds with high intensity values of the pixels.
f_7	Sum Variance	$$\sum_{n=2}^{2N} (n - f_6)^2 p_{x+y}(n)$$	Measures the variance of the intensity values. A high Sum Variance indicates a high variability in intensity values of the texture.
f_8	Sum Entropy	$$-\sum_{n=2}^{2N} p_{x+y}(n) \log(p_{x+y}(n))$$	Measures the randomness of the intensity values.
f_9	Entropy	$$-\sum_{i}\sum_{j} p(i,j) \log(p(i,j))$$	Measures the randomness of the distribution. A high Entropy indicates randomly distribution of the gray-levels
f_{10}	Difference Variance	$$\sum_{j}\left(j - \sum_{i} i p_{x-y}(i)\right)^2 p_{x-y}(j)$$	Measures the variance of the difference marginal probabilities.
f_{11}	Difference Entropy	$$-\sum_{n=0}^{N-1} p_{x-y}(n) \log(p_{x-y}(n))$$	Measures the entropy or amount of disorder within the difference marginal probabilities.
f_{12}	Information measure of correlation 1	$$\frac{f_9 + \sum_{i}\sum_{j} p(i,j) \log(p_x(i) p_x(j))}{-\sum_{i} p_x(i) log(p_x(i))}$$	Haralick lists two ways of measuring the information correlation.
f_{13}	Information measure of correlation 2	$$\sqrt{1 - e^{-2(H_{xy} - f_9)}}$$	

Table A.1: Haralick Texture features calculated from the Gray-Level Co-occurrence Matrix (GLCM).

A.2 Classifier Performance Measures

Notation	Name	Description
TP	True Positives	Number of correctly classified malignant samples
FP	False Positives	Number of incorrectly classified malignant samples
TN	True Negative	Number of correctly classified benign samples
FN	False Negative	Number of incorrectly classified benign samples
N		Number of overall samples

Performance Measure	Formula	Description
Sensitivity	$\dfrac{TP}{TP + FN}$	Sensitivity relates to the classifier's ability to detect positive(malignant) samples.
Specificity	$\dfrac{TN}{TN + FP}$	Specificity relates to the classifier's ability to identify negative (benign) results.
F-measure	$2 \times \dfrac{\text{Sensitivity} * \text{Specificity}}{\text{Sensitivity} + \text{Specificity}}$	Harmonic mean of precision and recall.
Correct Rate	$\dfrac{TP + TN}{N}$	Rate of correctly identified samples, both malignant and benign.
Error Rate	$\dfrac{FP + FN}{N}$	Rate of incorrectly identified samples.
Positive Predictive Value	$\dfrac{TP}{TP + FP}$	Proportion of samples with positive test results that are correctly identified.
Negative Predictive Value	$\dfrac{TN}{TN + FN}$	Proportion of samples with negative test results that are correctly identified.
AuC	$\displaystyle\int_0^1 ROC$	Area under ROC curve, where ROC is a plot of Sensitivity vs. (1 - Specificity) [112]

Table A.2: Summary of common performance measures for classifiers.

A.3 Classification Results

Features	F-Measure		Sensitivity		Specificity		Pos. Pred. Value		Neg. Pred. Value		Correct Rate		Error Rate	
	man	seg	man	seg	man	seg	man	seg	man	seg	man	seg	man	seg
$f_{l\text{-}adc} + f_{l\text{-}pet}$	0.8259	0.7544	0.8950	0.8550	0.7667	0.6750	0.8647	0.8143	0.8142	0.7364	0.8469	0.7875	0.1531	0.2125
$f_{l\text{-}adc}$	0.8163	0.8163	0.8000	0.8000	0.8333	0.8333	0.8889	0.8889	0.7143	0.7143	0.8125	0.8125	0.1875	0.1875
$f_{l\text{-}adc} + f_{l\text{-}\Delta ckc}$	0.7283	0.2756	0.8150	0.7950	0.6583	0.1667	0.7990	0.6139	0.6810	0.3279	0.7563	0.5594	0.2438	0.4406
$f_{l\text{-}adc} + f_{l\text{-}curve}$	0.6955	0.5926	0.9000	0.9000	0.5667	0.4417	0.7759	0.7287	0.7727	0.7260	0.7750	0.7281	0.2250	0.2719
$f_{l\text{-}adc} + f_{l\text{-}ckc} + f_{l\text{-}\Delta ckc}$	0.6884	0.3111	0.9200	0.8250	0.5500	0.1917	0.7731	0.6298	0.8049	0.3966	0.7813	0.5875	0.2188	0.4125
$f_{l\text{-}pet} + f_{l\text{-}adc} + f_{l\text{-}curve}$	0.6828	0.6632	0.9000	0.9000	0.5500	0.5250	0.7692	0.7595	0.7674	0.7590	0.7688	0.7594	0.2313	0.2406
$f_{l\text{-}\Delta ckc}$	0.6594	0.0165	0.8050	0.7450	0.5583	0.0083	0.7523	0.5560	0.6321	0.0192	0.7125	0.4688	0.2875	0.5313
$f_{l\text{-}ckc} + f_{l\text{-}\Delta ckc}$	0.6586	0.3309	0.9350	0.7000	0.5083	0.2167	0.7602	0.5983	0.8243	0.3023	0.7750	0.5188	0.2250	0.4813
$f_{l\text{-}adc} + f_{l\text{-}ckc}$	0.6279	0.3613	0.8000	0.8000	0.5167	0.2333	0.7339	0.6349	0.6078	0.4118	0.6938	0.5875	0.3063	0.4125
$f_{l\text{-}pet} + f_{l\text{-}curve}$	0.6269	0.6146	0.8400	0.9000	0.5000	0.4667	0.7368	0.7377	0.6522	0.7368	0.7125	0.7375	0.2875	0.2625
$f_{l\text{-}ckc}$	0.6154	0.4557	0.8000	0.7200	0.5000	0.3333	0.7273	0.6429	0.6000	0.4167	0.6875	0.5750	0.3125	0.4250
$f_{l\text{-}pet} + f_{l\text{-}ckc}$	0.5984	0.4688	0.7450	0.7900	0.5000	0.3333	0.7129	0.6639	0.5405	0.4878	0.6531	0.6188	0.3469	0.3813
$f_{l\text{-}curve}$	0.5691	0.4865	0.8000	0.9000	0.4417	0.3333	0.7049	0.6923	0.5699	0.6667	0.6656	0.6875	0.3344	0.3125
$f_{l\text{-}pet}$	0.2692	0.2692	0.7000	0.7000	0.1667	0.1667	0.5833	0.5833	0.2500	0.2500	0.5000	0.5000	0.5000	0.5000

Table A.3: Classification performance results using LOOCV with 20 iterations, sorted by F-value. Performance values are listed for the manual annotated lesions (*man*) and the segmented lesions (*seg*). The feature subset contains: DWI ADC value from fuzzy c-means clustering ($f_{l\text{-}adc}$), PET values from fuzzy c-means clustering ($f_{l\text{-}pet}$), data values from the CKC ($f_{l\text{-}ckc}$) and their derivatives ($f_{l\text{-}\Delta ckc}$), features from the fitted asymmetric logistic function ($f_{l\text{-}curve}$).

Features	F-Measure		Sensitivity		Specificity		Pos. Pred. Value		Neg. Pred. Value		Correct Rate		Error Rate	
	man	seg	man	seg	man	seg	man	seg	man	seg	man	seg	man	seg
$f_{l\text{-}adc}$	0.9091	0.9091	1.0000	1.0000	0.8333	0.8333	0.9091	0.9091	1.0000	1.0000	0.9375	0.9375	0.0625	0.0625
$f_{l\text{-}adc} + f_{l\text{-}\Delta ckc}$	0.8889	0.7270	0.8000	0.6550	1.0000	0.8167	1.0000	0.8562	0.7500	0.5868	0.8750	0.7156	0.1250	0.2844
$f_{l\text{-}adc} + f_{l\text{-}pet}$	0.8723	0.8241	0.9150	0.8150	0.8333	0.8333	0.9015	0.8907	0.8547	0.7299	0.8844	0.8219	0.1156	0.1781
$f_{l\text{-}adc} + f_{l\text{-}curve}$	0.8564	0.9269	0.8100	0.8700	0.9083	0.9917	0.9364	0.9943	0.7415	0.8207	0.8469	0.9156	0.1531	0.0844
$f_{l\text{-}\Delta ckc}$	0.8414	0.5990	0.7550	0.5750	0.9500	0.6250	0.9618	0.7188	0.6994	0.4688	0.8281	0.5938	0.1719	0.4063
$f_{l\text{-}adc} + f_{l\text{-}ckc}$	0.8111	0.6976	0.7900	0.6400	0.8333	0.7667	0.8876	0.8205	0.7042	0.5610	0.8063	0.6875	0.1938	0.3125
$f_{l\text{-}pet} + f_{l\text{-}adc} + f_{l\text{-}curve}$	0.8104	0.9381	0.8300	0.8900	0.7917	0.9917	0.8691	0.9944	0.7364	0.8440	0.8156	0.9281	0.1844	0.0719
$f_{l\text{-}adc} + f_{l\text{-}ckc} + f_{l\text{-}\Delta ckc}$	0.8000	0.5812	0.8800	0.4850	0.7333	0.7250	0.8462	0.7462	0.7857	0.4579	0.8250	0.5750	0.1750	0.4250
$f_{l\text{-}ckc} + f_{l\text{-}\Delta ckc}$	0.7647	0.5233	0.7800	0.5300	0.7500	0.5167	0.8387	0.6463	0.6716	0.3974	0.7688	0.5250	0.2313	0.4750
$f_{l\text{-}pet}$	0.7549	0.7609	0.6900	0.7000	0.8333	0.8333	0.8734	0.8750	0.6173	0.6250	0.7438	0.7500	0.2563	0.2500
$f_{l\text{-}ckc}$	0.7430	0.5484	0.7000	0.6200	0.7917	0.4917	0.8485	0.6703	0.6129	0.4370	0.7344	0.5719	0.2656	0.4281
$f_{l\text{-}curve}$	0.7395	0.6446	0.6700	0.6100	0.8250	0.6833	0.8645	0.7625	0.6000	0.5125	0.7281	0.6375	0.2719	0.3625
$f_{l\text{-}pet} + f_{l\text{-}ckc}$	0.7176	0.5357	0.6950	0.7500	0.7417	0.4167	0.8177	0.6818	0.5933	0.5000	0.7125	0.6250	0.2875	0.3750
$f_{l\text{-}pet} + f_{l\text{-}curve}$	0.7034	0.7636	0.7550	0.8400	0.6583	0.7000	0.7865	0.8235	0.6172	0.7241	0.7188	0.7875	0.2813	0.2125

Table A.4: Classification performance results using LOOCV with 20 iterations using an optimized threshold value for the Random Forest prediction. The table is sorted by F-value.

Bibliography

[1] R. Adams and L. Bischof. Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6):641–647, 1994.

[2] S. Agner, S. Soman, E. Libfeld, M. McDonald, K. Thomas, S. Englander, M. Rosen, D. Chin, J. Nosher, and A. Madabhushi. Textural kinetics: A novel Dynamic Contrast-Enhanced (DCE)-MRI feature for breast lesion classification. *Journal of Digital Imaging*, 24:446–463, 2011.

[3] S. Agner, J. Xu, M. Rosen, S. Karthigeyan, S. Englander, and A. Madabhushi. Spectral embedding based active contour (SEAC): application to breast lesion segmentation on DCE-MRI. In *SPIE Medical Imaging*, volume 7963, pages 796305.1 – 796305.12, 2011.

[4] D. Aloise, A. Deshpande, P. Hansen, and P. Popat. Np-hardness of euclidean sum-of-squares clustering. *Machine Learning*, 75:245–248, 2009.

[5] G. Antoch and A. Bockisch. Combined PET/MRI: a new dimension in whole-body oncology imaging? *European Journal of Nuclear Medicine and Molecular Imaging*, 36:113–120, 2009.

[6] K. J. Archer and R. V. Kimes. Empirical characterization of random forest variable importance measures. *Computational Statistics & Data Analysis*, 52(4):2249 – 2260, 2008.

[7] E. Ardizzone, O. Gambino, M. La Cascia, L. Presti, and R. Pirrone. Multi-modal non-rigid registration of medical images based on mutual information maximization. In *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on*, pages 743–750. IEEE, 2007.

[8] V. Arsigny, O. Commowick, X. Pennec, and N. Ayache. A log-Euclidean framework for statistics on diffeomorphisms. *Med Image Comput Comput Assist Interv*, 9:924–931, 2006.

[9] J. Ashburner. A fast diffeomorphic image registration algorithm. *Neuroimage*, 38:95–113, Oct 2007.

[10] J. Ashburner and K. Friston. Rigid body registration. In R. Frackowiak, K. Friston, C. Frith, R. Dolan, K. Friston, C. Price, S. Zeki, J. Ashburner, and W. Penny, editors, *Human Brain Function*. Academic Press, 2nd edition, 2003.

[11] J. Ashburner and K. J. Friston. Unified segmentation. *NeuroImage*, 26(3):839 – 851, 2005.

[12] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal*, 12(1):26–41, 2 2008.

[13] B. B. Avants, P. T. Schoenemann, and J. C. Gee. Lagrangian frame diffeomorphic image registration: Morphometric comparison of human and chimpanzee cortex. *Medical Image Analysis*, 10(3):397 – 412, 2006.

[14] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee. A reproducible evaluation of ants similarity metric performance in brain image registration. *NeuroImage*, 54(3):2033 – 2044, 2011.

[15] T. Ayer, M. U. Ayvaci, Z. X. Liu, O. Alagoz, and E. S. Burnside. Computer-aided diagnostic models in breast cancer screening. *Imaging Med*, 2:313–323, Jun 2010.

[16] R. Bammer. Basic principles of diffusion-weighted imaging. *European Journal of Radiology*, 45(3):169 – 184, 2003.

[17] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12:43–77, 1994.

[18] M. Baum. The curability of breast cancer. *Br Med J.*, 1(6007):439–442, February 1976.

[19] M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision*, 61(2):139–157, 2005.

[20] W. Beil, K. Rohr, and H. S. Stiehl. Investigation of approaches for the localization of anatomical landmarks in 3d medical images. In H. Lemke, M. Vannier, and K. Inamura, editors, *Proc. Computer Assisted Radiology and Surgery*, pages 265–270. Elsevier Amsterdam Lausanne, 1997.

[21] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Kluwer Academic Publishers, Norwell, MA, USA, 1981.

[22] J. C. Bezdek, R. Ehrlich, and W. Full. FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2–3):191 – 203, 1984.

[23] W. Bidgood Jr., S. Horii, F. Prior, and D. Van Syckle. Understanding and using DICOM, the data interchange standard for biomedical imaging. *Journal of the American Medical Informatics Association*, 4(3):199–212, 1997.

[24] W. Bogner, S. Gruber, K. Pinker, G. Grabner, A. Stadlbauer, M. Weber, E. Moser, T. H. Helbich, and S. Trattnig. Diffusion-weighted MR for differentiation of breast lesions at 3.0 T: How does selection of diffusion protocols affect diagnosis? *Radiology*, 253(2):341–351, 2009.

[25] W. Bogner, K. Pinker-Domenig, H. Bickel, M. Chmelik, M. Weber, T. H. Helbich, S. Trattnig, and S. Gruber. Readout-segmented echo-planar imaging improves the diagnostic performance of Diffusion-weighted MR breast examinations at 3.0 T. *Radiology*, 263(1):64–76, Apr 2012.

[26] J. Bonnans. *Numerical optimization: theoretical and practical aspects*. Springer, 2006.

[27] F. L. Bookstein. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, 1989.

[28] L. Breiman. Random forests. *Machine Learning*, 45:5–32, 2001.

[29] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen. *Classification and Regression Trees*. Chapman & Hall/CRC, 1 edition, Jan. 1984.

[30] L. G. Brown. A survey of image registration techniques. *ACM computing surveys (CSUR)*, 24(4):325–376, 1992.

[31] W. Buntine and T. Niblett. A further comparison of splitting rules for decision-tree induction. *Machine Learning*, 8:75–85, 1992.

108

[32] P. Cachier, X. Pennec, and N. Ayache. Fast non rigid matching by gradient descent: Study and improvements of the Demons algorithm. Technical Report RR-3706, INRIA, Jun 1999.

[33] L. B. Chao Chen, Andy Liaw. Using random forest to learn imbalanced data. Technical report, Department of Statistics,UC Berkeley, 2004.

[34] C. Chefd'hotel, G. Hermosillo, and O. Faugeras. Flows of diffeomorphisms for multimodal image registration. In *IEEE International Symposium on Biomedical Imaging. Proceedings.*, pages 753 – 756, 2002.

[35] 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. *Academic Radiology*, 13(1):63 – 72, 2006.

[36] 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. *Medical Physics*, 33(8):2878–87, 2006.

[37] W. Chen, M. L. Giger, H. Li, U. Bick, and G. M. Newstead. Volumetric texture analysis of breast lesions on contrast-enhanced magnetic resonance images. *Magnetic Resonance in Medicine*, 58(3):562–571, 2007.

[38] Z.-H. Cho, Y.-D. Son, H.-K. Kim, K.-N. Kim, S.-H. Oh, J.-Y. Han, I.-K. Hong, and Y.-B. Kim. A fusion PET-MRI system with a high-resolution research tomograph-PET and ultra-high field 7.0 T-MRI for the molecular-genetic imaging of the brain. *PROTEOMICS*, 8(6):1302–1323, 2008.

[39] Z.-H. Cho, Y.-D. Son, Y.-B. Kim, and S.-S. Yoo. Fusion of PET and MRI for hybrid imaging. In T. M. Deserno, editor, *Biomedical Image Processing*, Biological and Medical Physics, Biomedical Engineering, pages 55–79. Springer Berlin Heidelberg, 2011.

[40] G. Christensen, R. Rabbitt, and M. Miller. Deformable templates using large deformation kinematics. *IEEE Transactions on Image Processing*, 5(10):1435 –1447, oct 1996.

[41] E. Coddington and N. Levinson. *Theory of ordinary differential equations.* International series in pure and applied mathematics. McGraw-Hill, 1955.

[42] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multi-modality image registration based on information theory. *Information Processing in Medical Imaging*, pages 263–274, 1995.

[43] W. R. Crum, T. Hartkens, and D. L. G. Hill. Non-rigid image registration: theory and practice. *British journal of radiology*, 77(Special Issue 2):S140, 2004.

[44] T. Deserno. *Biomedical Image Processing*. Biological and Medical Physics, Biomedical Engineering. Springer, 2011.

[45] L. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26 (3):297–302, 1945.

[46] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez. Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence*, 89(1–2):31 – 71, 1997.

[47] Q. Du, V. Faber, and M. Gunzburger. Centroidal voronoi tessellations: Applications and algorithms. *SIAM Review*, 41(4):637–676, 1999.

[48] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley, 2001.

[49] P. Dupuis and U. Grenander. Variational problems on flows of diffeomorphisms for image matching. *Q. Appl. Math.*, LVI:587–600, September 1998.

[50] D. G. Ebin and J. Marsden. Groups of diffeomorphisms and the motion of an incompressible fluid.groups of diffeomorphisms and the motion of an incompressible fluid. *Annals of Mathematics*, 92(1):102–163, 1970.

[51] E. Eyal and H. Degani. Model-based and model-free parametric analysis of breast dynamic-contrast-enhanced MRI. *NMR in Biomedicine*, 22(1):40–53, 2009.

[52] S. Feig. Breast masses. mammographic and sonographic evaluation. *Radiol Clin North Am.*, 30(1):67–92, 1992.

[53] J. Ferlay, H.-R. Shin, F. Bray, D. Forman, C. Mathers, and D. M. Parkin. Estimates of worldwide burden of cancer in 2008: Globocan 2008. *International journal of cancer*, 127(12):2893–2917, 2010.

[54] D. R. Forsey and R. H. Bartels. Hierarchical b-spline refinement. *SIGGRAPH Comput. Graph.*, 22:205–212, June 1988.

[55] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

[56] B. Fueger, T. Helbich, M. Schernthaner, S. Zbyn, H.-G. Linhart, A. Stiglbauer, A. Doan, K. Pinker, G. Heinz, A. Padhani, and P. Brader. Importance of multiparametric magnetic resonance tomography for prostate cancer. *Der Radiologe*, 51:947–954, 2011.

[57] J. C. Gee, , M. Reivich, and R. Bajcsy. Elastically deforming a three-dimensional atlas to match anatomical brain images. *J. Comput. Assist. Tomogr*, 17:225–236, 1993.

[58] J. C. Gee. On matching brain volumes. *Pattern Recognition*, 32:99–111, 1998.

[59] R. Genuer, J.-M. Poggi, and C. Tuleau-Malot. Variable selection using random forests. *Pattern Recognition Letters*, 31(14):2225 – 2236, 2010.

[60] P. Gibbs and L. W. Turnbull. Textural analysis of contrast-enhanced MR images of the breast. *Magnetic Resonance in Medicine*, 50(1):92–98, 2003.

[61] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2001.

[62] A. Goshtasby. Registration of images with geometric distortions. *IEEE Transactions on Geoscience and Remote Sensing*, 26:60–64, Jan. 1988.

[63] S. Gruber, B.-K. Debski, K. Pinker, M. Chmelik, G. Grabner, T. Helbich, S. Trattnig, and W. Bogner. Three-dimensional proton MR spectroscopic imaging at 3 T for the differentiation of benign and malignant breast lesions. *Radiology*, 261:752–761, 2011.

[64] A. Guimond, A. Roche, N. Ayache, and J. Meunier. Multimodal brain warping using the demons algorithm and adaptative intensity corrections. Technical Report RR-3796, INRIA, Nov. 1999.

[65] Y. Guo, Y.-Q. Cai, Z.-L. Cai, Y.-G. Gao, N.-Y. An, L. Ma, S. Mahankali, and J.-H. Gao. Differentiation of clinically benign and malignant breast lesions using diffusion-weighted imaging. *Journal of Magnetic Resonance Imaging*, 16(2):172–178, 2002.

110

[66] Y. Guo, R. Sivaramakrishna, C.-C. Lu, J. Suri, and S. Laxminarayan. Breast image registration techniques: a survey. *Medical and Biological Engineering and Computing*, 44:15–26, 2006.

[67] D. Han, J. Bayouth, Q. Song, A. Taurani, M. Sonka, J. Buatti, and X. Wu. Globally optimal tumor segmentation in PET-CT images: A graph-based co-segmentation method. In G. Székely and H. Hahn, editors, *Information Processing in Medical Imaging*, volume 6801 of *Lecture Notes in Computer Science*, pages 245–256. Springer Berlin / Heidelberg, 2011.

[68] R. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786 – 804, may 1979.

[69] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning, 2nd ed*. Springer, 2009.

[70] H. He and E. A. Garcia. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21:1263–1284, 2009.

[71] G. Hermosillo, C. Chefd'Hotel, and O. Faugeras. Variational methods for multimodal image matching. *International Journal of Computer Vision*, 50:329–343, 2002.

[72] M. Hernandez, M. N. Bossa, and S. Olmos. Registration of anatomical images using geodesic paths of diffeomorphisms parameterized with stationary vector fields. In *IEEE 11th International Conference on Computer Vision*, pages 1 –8, oct. 2007.

[73] A. Hill, A. Mehnert, S. Crozier, and K. McMahon. Evaluating the accuracy and impact of registration in dynamic contrast-enhanced breast mri. *Concepts in Magnetic Resonance Part B: Magnetic Resonance Engineering*, 35B(2):106–120, 2009.

[74] H. Huang, C. Ceritoglu, X. Li, A. Qiu, M. I. Miller, P. C. van Zijl, and S. Mori. Correction of B0 susceptibility induced distortion in diffusion-weighted images using large-deformation diffeomorphic metric mapping. *Magnetic Resonance Imaging*, 26(9):1294 – 1302, 2008.

[75] R. A. Hubbard, K. Kerlikowske, C. I. Flowers, B. C. Yankaskas, W. Zhu, and D. L. Miglioretti. Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study. *Ann. Intern. Med.*, 155:481–492, Oct 2011.

[76] L. Ibanez and W. Schroeder. *The ITK Software Guide 2.4*. Kitware, Inc., Nov. 2005.

[77] D. M. Ikeda, N. M. Hylton, K. Kinkel, M. G. Hochman, C. K. Kuhl, W. A. Kaiser, J. C. Weinreb, S. F. Smazal, H. Degani, P. Viehweg, J. Barclay, and M. D. Schnall. Development, standardization, and testing of a lexicon for reporting contrast-enhanced breast magnetic resonance imaging studies. *J Magn Reson Imaging*, 13:889–895, Jun 2001.

[78] M. Judenhofer, H. Wehrl, D. Newport, C. Catana, S. Siegel, M. Becker, A. Thielscher, M. Kneilling, M. Lichy, M. Eichner, K. Klingel, G. Reischl, S. Widmaier, M. Röcken, R. Nutt, H.-J. Machulla, K. Uludag, S. Cherry, C. Claussen, and B. Pichler. Simultaneous PET-MRI: a new approach for functional and morphological imaging. *Nature Medicine*, 14:459 – 465, 2008.

[79] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1:321–331, 1988.

[80] D. Keysers and W. Unger. Elastic image matching is NP-complete. *Pattern Recognition Letters*, 24(1–3):445 – 453, 2003.

111

[81] K. Kinkel, T. H. Helbich, L. J. Esserman, J. Barclay, E. H. Schwerin, E. A. Sickles, and N. M. Hylton. Dynamic high-spatial-resolution MR imaging of suspicious breast lesions: diagnostic criteria and interobserver variability. *AJR Am J Roentgenol*, 175:35–43, Jul 2000.

[82] A. Klein, J. Andersson, B. A. Ardekani, J. Ashburner, B. Avants, M.-C. 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*, 46(3):786–802, 7 2009.

[83] T. M. Kolb, J. Lichy, and J. H. Newhouse. Occult cancer in women with dense breasts: detection with screening US–diagnostic yield and tumor characteristics. *Radiology*, 207:191–199, Apr 1998.

[84] D.-J. Kroon and K. Slump. MRI modalitiy transformation in demon registration. In *ISBI'09*, pages 963–966, 2009.

[85] C. Kuhl. The current status of breast MR imaging part 1. choice of technique, image interpretation, diagnostic accuracy, and transfer to clinical practice. *Radiology*, 244(2):356–378, August 2007.

[86] C. K. Kuhl, P. Jost, N. Morakkabati, O. Zivanovic, H. H. Schild, and J. Gieseke. Contrast-enhanced MR imaging of the breast at 3.0 and 1.5 T in the same patients: Initial experience. *Radiology*, 239(3):666–676, June 2006.

[87] 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*, 211(1):101–110, 1999.

[88] Kunio and Doi. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Computerized Medical Imaging and Graphics*, 31(4–5):198 – 211, 2007.

[89] S. Lee, G. Wolberg, and S. Y. Shin. Scattered data interpolation with multilevel b-splines. *IEEE Transactions on Visualization and Computer Graphics*, 3:228–244, 1997.

[90] N. Lepore, C. Brun, Y. Y. Chou, M. C. Chiang, R. A. Dutton, K. M. Hayashi, E. Luders, O. L. Lopez, H. J. Aizenstein, A. W. Toga, J. T. Becker, and P. M. Thompson. Generalized tensor-based morphometry of HIV/AIDS using multivariate statistics on deformation tensors. *IEEE Trans Med Imaging*, 27:129–141, Jan 2008.

[91] J. Levman, T. Leung, P. Causer, D. Plewes, and A. Martel. Classification of dynamic contrast-enhanced magnetic resonance breast lesions by support vector machines. *IEEE Transactions on Medical Imaging*, 27(5):688 –696, may 2008.

[92] J. P. Lewis. Fast normalized cross-correlation. In *Vision Interface*, pages 120–123. Canadian Image Processing and Pattern Recognition Society, 1995.

[93] S. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129 – 137, mar 1982.

[94] D. Lowe. Object recognition from local scale-invariant features. In *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, volume 2, pages 1150 –1157 vol.2, 1999.

[95] H. Lu, M. Reyes, A. Šerijović, S. Weber, Y. Sakurai, H. Yamagata, and P. C. Cattin. Multi-modal diffeomorphic demons registration based on point-wise mutual information. In *Proceedings of the 2010 IEEE international conference on Biomedical imaging: from nano to Macro*, ISBI'10, pages 372–375, Piscataway, NJ, USA, 2010. IEEE Press.

[96] F. Maes, A. Collignon, D. V, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE transactions on Medical Imaging*, 16:187–198, 1997.

[97] C. Marini, C. Iacconi, M. Giannelli, A. Cilotti, M. Moretti, and C. Bartolozzi. Quantitative diffusion-weighted MR imaging in the differential diagnosis of breast lesion. *European Radiology*, 17:2646–2655, 2007.

[98] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank. PET-CT image registration in the chest using free-form deformations. *IEEE Trans Med Imaging*, 22:120–128, Jan 2003.

[99] R. McGill, J. W. Tukey, and W. A. Larsen. Variations of box plots. *The American Statistician*, 32:12–16, 1978.

[100] C. E. McLaren, W.-P. Chen, K. Nie, and M.-Y. Su. Prediction of malignant breast lesions from MRI features: A comparison of artificial neural network and logistic regression techniques. *Academic Radiology*, 16(7):842 – 851, 2009.

[101] L. A. Meinel, A. H. Stolpen, K. S. Berbaum, L. L. Fajardo, and J. M. Reinhardt. Breast MRI lesion classification: Improved performance of human readers with a backpropagation neural network computer-aided diagnosis (CAD) system. *Journal of Magnetic Resonance Imaging*, 25(1):89–95, 2007.

[102] M. I. Miga. A new approach to elastography using mutual information and finite elements. *Physics in Medicine and Biology*, 48(4):467, 2003.

[103] M. Miller, A. Trouve, and L. Younes. On the metrics and Euler-Lagrange equations of computational anatomy. *Annual Review of Biomedical Engineering*, 4(1):375–405, 2002.

[104] M. I. Miller. Computational anatomy: shape, growth, and atrophy comparison via diffeomorphisms. *Neuroimage*, 23 Suppl 1:19–33, 2004.

[105] P. Mitteroecker and P. Gunz. Advances in geometric morphometrics. *Evolutionary Biology*, 36:235–247, 2009.

[106] M. Modat, T. Vercauteren, G. R. Ridgway, D. J. Hawkes, N. C. Fox, and S. Ourselin. Diffeomorphic demons using normalized mutual information, evaluation on multimodal brain MR images. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7623, 2010.

[107] M. Moskowitz. *Cancer of the Breast*, chapter Breast imaging, pages 206–239. Philadelphia, PA.: W B Saunders, 1995.

[108] L. Moy, F. Ponzo, M. E. Noz, G. Q. Maguire Jr, A. D. Murphy-Walcott, A. E. Deans, M. T. Kitazono, L. Travascio, and E. L. Kramer. Improving specificity of breast MRI using prone PET and fused MRI and PET 3D volume datasets. *Journal of Nuclear Medicine*, 48(4):528, 2007.

[109] K. Murphy, B. van Ginneken, J. M. Reinhardt, S. Kabus, K. Ding, X. Deng, K. Cao, K. Du, G. E. Christensen, V. Garcia, T. Vercauteren, N. Ayache, O. Commowick, G. Malandain, B. Glocker, N. Paragios, N. Navab, V. Gorbunova, J. Sporring, M. de Bruijne, X. Han, M. P. Heinrich, J. A. Schnabel, M. Jenkinson, C. Lorenz, M. Modat, J. R. McClelland, S. Ourselin, S. E. Muenzing, M. A. Viergever, D. De Nigris, D. L. Collins, T. Arbel, M. Peroni, R. Li, G. C. Sharp, A. Schmidt-Richberg, J. Ehrhardt, R. Werner, D. Smeets, D. Loeckx, G. Song, N. Tustison, B. Avants, J. C. Gee, M. Staring, S. Klein, B. C. Stoel, M. Urschler, M. Werlberger, J. Vandemeulebroucke, S. Rit, D. Sarrut, and J. P. Pluim. Evaluation of registration methods on thoracic CT: the EMPIRE10 challenge. *IEEE Trans Med Imaging*, 30:1901–1920, Nov 2011.

113

[110] S. Mussurakis, D. L. Buckley, and A. Horsman. Dynamic MRI of invasive breast cancer: assessment of three region-of-interest analysis methods. *J Comput Assist Tomogr*, 21:431–438, 1997.

[111] K. Nie, J.-H. Chen, H. J. Yu, Y. Chu, O. Nalcioglu, and M.-Y. Su. Quantitative analysis of lesion morphology and texture features for diagnostic prediction in breast MRI. *Academic Radiology*, 15(12):1513 – 1525, 2008.

[112] N. A. Obuchowski. Receiver operating characteristic curves and their use in radiology. *Radiology*, 229:3–8, Oct 2003.

[113] N. Otsu. A threshold selection method from graylevel histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9:62–66, 1979.

[114] E. Parzen. On the estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33:1065–1076, 1962.

[115] H. Peng, F. Long, and C. Ding. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:1226–1238, 2005.

[116] F. P. A. Pereira, G. Martins, and R. de Vasconcellos Carvalhaes de Oliveira. Diffusion magnetic resonance imaging of the breast. *Magnetic Resonance Imaging Clinics of North America*, 19(1):95 – 110, 2011.

[117] K. Pinker, H. Bickel, H. Magometschnigg, B. Brück, S. Gruber, W. Bogner, P. Brader, and T. H. Helbich. Molecular imaging of breast tumors with PET-MRI: proof of concept. In *ECR Vienna*, 2011.

[118] K. Pinker, P. Brader, G. Karanikas, K. El-Rabadi, W. Bogner, S. Gruber, M. Reisegger, S. Trattnig, and T. Helbich. Functional and molecular imaging of breast tumors. *Radiologe*, 50(11):1030–1038, November 2010.

[119] K. Pinker, G. Grabner, W. Bogner, S. Gruber, S. Trattnig, G. Heinz-Peer, F. Fitzal, U. Pluschnig, M. Rudas, and T. Helbich. A combined high temporal and high spatial resolution 3 tesla mr imaging protocol for the assessment of breast lesions: initial results. *Investigative Radiology*, 44(9):553–558, September 2009.

[120] K. Pinker, G. Karanikas, M. Mayerhoefer, K. El-Rabadi, S. M., and H. TH. Molecular imaging in breast lesions: comparison of dedicated breast 18F-FDG-PET-CT vs. contrast-enhanced MRI at 3 Tesla. In *ECR Vienna*, 2010.

[121] W. Plishker, O. Dandekar, S. Bhattacharyya, and R. Shekhar. A taxonomy for medical image registration acceleration techniques. In *Life Science Systems and Applications Workshop, 2007. LISA 2007. IEEE/NIH*, pages 160 –163, nov. 2007.

[122] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual-information-based registration of medical images: a survey. *IEEE Trans Med Imaging*, 22(8):986–1004, 8 2003.

[123] D. A. Porter and R. M. Heidemann. High resolution diffusion-weighted imaging using readout-segmented echo-planar imaging, parallel imaging and a two-dimensional navigator-based reacquisition. *Magnetic Resonance in Medicine*, 62(2):468–475, 2009.

[124] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge, UK, 2nd edition, 1992.

114

[125] J. Quinlan. *C4.5: programs for machine learning*. Morgan Kaufmann series in machine learning. Morgan Kaufmann Publishers, 1993.

[126] C. S. Reiner, M. A. Fischer, T. Hany, P. Stolzmann, D. Nanz, O. F. Donati, D. Weishaupt, G. K. von Schulthess, and H. Scheffel. Molecular imaging of malignant tumor metabolism: Whole-body image fusion of DWI/CT vs. PET/CT. *Academic Radiology*, 18(8):940 – 946, 2011.

[127] X. Ren and J. Malik. Learning a classification model for segmentation. In *Ninth IEEE International Conference on Computer Vision, Proceedings*, pages 10 –17 vol.1, oct. 2003.

[128] F. J. Richards. A flexible growth function for empirical use. *Journal of Experimental Botany*, 10(2):290–301, 1959.

[129] T. Rohlfing, C. R. M. Jr., D. A. Bluemke, and M. A. Jacobs. Volume-preserving nonrigid registration of MR breast images using free-form deformation with an incompressibility constraint. *IEEE Transactions on Medical Imaging*, 22:730–741, 2003.

[130] K. Rohr. *Landmark-Based Image Analysis: Using Geometric and Intensity Models*. Kluwer Academic Publishers, Norwell, MA, USA, 2001.

[131] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837, 1956.

[132] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imaging*, 18(8):712–21, 8 1999.

[133] Y. Saeys, T. Abeel, and Y. Van de Peer. Robust feature selection using ensemble feature selection techniques. In W. Daelemans, B. Goethals, and K. Morik, editors, *Machine Learning and Knowledge Discovery in Databases*, volume 5212 of *Lecture Notes in Computer Science*, pages 313–325. Springer Berlin / Heidelberg, 2008.

[134] B. P. Schneider and K. D. Miller. Angiogenesis of breast cancer. *J. Clin. Oncol.*, 23:1782–1790, Mar 2005.

[135] L. A. Schwarz. Non-rigid registration using free-form deformations. Master's thesis, Technische Universität München, Fakultät für Informatik, May 2007.

[136] T. W. Sederberg and S. R. Parry. Free-form deformation of solid geometric models. *SIGGRAPH Comput. Graph.*, 20:151–160, August 1986.

[137] Z. Y. Shan, S. J. Mateja, W. E. Reddick, J. O. Glass, and B. L. Shulkin. Retrospective evaluation of PET-MRI registration algorithms. *J Digit Imaging*, 24:485–493, Jun 2011.

[138] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423/623–656, 1948.

[139] J. Shi, B. Sahiner, H. P. Chan, C. Paramagul, L. M. Hadjiiski, M. Helvie, and T. Chenevert. Treatment response assessment of breast masses on dynamic contrast-enhanced magnetic resonance scans using fuzzy c-means clustering and level set segmentation. *Med Phys*, 36:5052–5063, Nov 2009.

[140] E. J. Somer, N. A. Benatar, M. J. O'Doherty, M. A. Smith, and P. K. Marsden. Use of the CT component of PET–CT to improve PET–MR registration: demonstration in soft-tissue sarcoma. *Physics in Medicine and Biology*, 52(23):6991, 2007.

[141] C. Strobl, A. L. Boulesteix, A. Zeileis, and T. Hothorn. Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics*, 8:25, 2007.

[142] T. Strutz. *Data Fitting and Uncertainty: A Practical Introduction to Weighted Least Squares and Beyond*. Vieweg and Teubner, Germany, 2010.

[143] C. Studholme, C. Drapaca, B. Iordanova, and V. Cardenas. Deformation-based mapping of volume change from serial brain MRI in the presence of local tissue contrast change. *IEEE Trans Med Imaging*, 25:626–639, 2006.

[144] C. Studholme, D. L. G. Hill, and D. J. Hawkes. An overlap invariant entropy measure of 3d medical image alignment. *Pattern Recognition*, 32(1):71–86, Jan. 1999.

[145] B. K. Szabo, M. K. Wiberg, B. Bone, and P. Aspelin. Application of artificial neural networks to the analysis of dynamic MR imaging features of the breast. *Eur Radiol*, 14:1217–1225, Jul 2004.

[146] C. Tanner, J. Schnabel, D. Chung, M. Clarkson, D. Rueckert, D. Hill, 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 S. Delp, A. DiGoia, and B. Jaramaz, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2000*, volume 1935 of *Lecture Notes in Computer Science*, pages CH371–CH371. Springer Berlin / Heidelberg, 2000.

[147] M. Tatsumi, C. Cohade, K. Mourtzikos, E. Fishman, and R. Wahl. Initial experience with FDG-PET/CT in the evaluation of breast cancer. *European Journal of Nuclear Medicine and Molecular Imaging*, 33:254–262, 2006.

[148] J.-P. Thirion. Image matching as a diffusion process: an analogy with Maxwell's demons. *Medical Image Analysis*, 2(3):243 – 260, 1998.

[149] J.-P. Thirion and A. Gourdon. Computing the differential characteristics of isointensity surfaces. *Computer Vision and Image Understanding*, 61(2):190 – 202, 1995.

[150] P. Thompson and A. Toga. A surface-based technique for warping three-dimensional images of the brain. *IEEE Transactions on Medical Imaging*, 15(4):402 –417, aug 1996.

[151] A. Trouve. Diffeomorphisms Groups and Pattern Matching in Image Analysis. *International Journal of Computer Vision*, 28(3):213–221, July 1998.

[152] T. Twellmann, O. Lichte, and T. Nattkemper. An adaptive tissue characterization network for model-free visualization of dynamic contrast-enhanced magnetic resonance image data. *IEEE Transactions on Medical Imaging*, 24(10):1256 –1266, oct. 2005.

[153] M. Unlu, A. Krol, A. Magri, J. Mandel, W. Lee, K. Baum, E. Lipson, I. Coman, and D. Feiglin. Computerized method for nonrigid MR-to-PET breast-image registration. *Computers in Biology and Medicine*, 40(1):37 – 53, 2010.

[154] T. Vercauteren, X. Pennec, E. Malis, A. Perchant, and N. Ayache. Insight into efficient image registration techniques and the demons algorithm. In N. Karssemeijer and B. Lelieveldt, editors, *Information Processing in Medical Imaging*, volume 4584 of *Lecture Notes in Computer Science*, pages 495–506. Springer Berlin / Heidelberg, 2007.

[155] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. Non-parametric diffeomorphic image registration with the demons algorithm. *Med Image Comput Comput Assist Interv*, 10(Pt 2):319–26, 2007.

[156] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. Diffeomorphic demons: efficient non-parametric image registration. *Neuroimage*, 45(1 Suppl):S61–72, 3 2009.

[157] A. Vignati, V. Giannini, A. Bert, M. Deluca, L. Morra, D. Persano, L. Martincich, and D. Regge. A fully automatic lesion detection method for DCE-MRI fat-suppressed breast images. In N. Karssemeijer and M. L. Giger, editors, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7260, pages 726026.1–726026.12. SPIE, feb 2009.

[158] P. Viola and W. M. Wells III. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24:137–154, 1997.

[159] C. J. Vyborny, M. L. Giger, and R. M. Nishikawa. Computer-aided diagnosis and detection of breast cancer. *Radiologic Clinics of North America*, 38(4):725 – 740, 2000.

[160] G. Wahba. *Spline Models for Observational Data (CBMS-NSF Regional Conference Series in Applied Mathematics)*. SIAM: Society for Industrial and Applied Mathematics (September 1, 1990), 1990.

[161] H. Wang, L. Dong, J. O'Daniel, R. Mohan, A. S. Garden, K. K. Ang, D. A. Kuban, M. Bonnen, J. Y. Chang, and R. Cheung. Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy. *Phys Med Biol*, 50:2887–2905, Jun 2005.

[162] M. Y. Wang, C. R. Maurer, J. M. Fitzpatrick, and R. J. Maciunas. An automatic technique for finding and localizing externally attached markers in CT and MR volume images of the head. *IEEE Trans Biomed Eng*, 43:627–637, Jun 1996.

[163] E. Warner, D. B. Plewes, R. S. Shumak, G. C. Catzavelos, L. S. Di Prospero, M. J. Yaffe, V. Goel, E. Ramsay, P. L. Chart, D. E. Cole, G. A. Taylor, M. Cutrara, T. H. Samuels, J. P. Murphy, J. M. Murphy, and S. A. Narod. Comparison of breast magnetic resonance imaging, mammography, and ultrasound for surveillance of women at high risk for hereditary breast cancer. *Journal of Clinical Oncology*, 19(15):3524–3531, 2001.

[164] J. C. Weinreb and G. Newstead. MR imaging of the breast. *Radiology*, 196(3):593–610, 1995.

[165] W. M. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35 – 51, 1996.

[166] M. A. Wirth, J. Narhan, and D. Gray. A model for nonrigid mammogram registration using mutual information. In *Digital mammography IWDM*, page 243–245, Bremen, Germany, 2002.

[167] G. Wolberg. *Digital Image Warping*. IEEE Computer Society Press, Los Alamitos, CA, USA, 1st edition, 1994.

[168] B. J. Woods. *Computer-aided detection of malignant lesions in dynamic contrast enhanced MRI breast and prostate cancer datasets*. PhD thesis, The Ohio State University, 2008.

[169] B. J. Woods, B. D. Clymer, T. Kurc, J. T. Heverhagen, R. Stevens, A. Orsdemir, O. Bulan, and M. V. Knopp. Malignant-lesion segmentation using 4d co-occurrence texture analysis applied to dynamic contrast-enhanced magnetic resonance breast image data. *Journal of Magnetic Resonance Imaging*, 25(3):495–501, 2007.

[170] Q. Wu, M. Salganicoff, A. Krishnan, D. S. Fussell, and M. K. Markey. Interactive lesion segmentation on dynamic contrast enhanced breast MRI using a markov model. In J. M. R. . J. P. W. Pluim, editor, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 6144, pages 1487–1494, Mar. 2006.

[171] H. Yabuuchi, Y. Matsuo, T. Kamitani, T. Setoguchi, T. Okafuji, H. Soeda, S. Sakai, M. Hatakenaka, M. Kubo, E. Tokunaga, H. Yamamoto, and H. Honda. Non-mass-like enhancement on contrast-enhanced breast MR imaging: lesion characterization using combination of dynamic contrast-enhanced and diffusion-weighted MR images. *Eur J Radiol*, 75:e126–132, Jul 2010.

[172] H. Yabuuchi, Y. Matsuo, T. Okafuji, T. Kamitani, H. Soeda, T. Setoguchi, S. Sakai, M. Hatakenaka, M. Kubo, N. Sadanaga, H. Yamamoto, and H. Honda. Enhanced mass on contrast-enhanced breast MR imaging: Lesion characterization using combination of dynamic contrast-enhanced and diffusion-weighted MR images. *J Magn Reson Imaging*, 28:1157–1165, Nov 2008.

[173] T. Yankeelov, M. Lepage, A. Chakravarthy, E. Broome, K. Niermann, M. Kelley, I. Meszoely, I. Mayer, C. Herman, K. McManus, R. Price, and J. Gore. Integration of quantitative DCE-MRI and ADC mapping to monitor treatment response in human breast cancer: initial results. *Magnetic Resonance Imaging*, 25(1):1–13, 2007.

[174] J. Yao, J. Chen, and C. Chow. Breast tumor analysis in dynamic contrast enhanced MRI using texture features and wavelet transform. *IEEE Journal of Selected Topics in Signal Processing*, 3:94–100, Feb. 2009.

[175] B. T. Yeo, T. Vercauteren, P. Fillard, X. Pennec, P. Golland, N. Ayache, and O. Clatz. DTI registration with exact finite-strain differential. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'08)*, pages 700–703, Paris, France, 2008. IEEE.

[176] F. F. Yin, M. L. Giger, K. Doi, C. J. Vyborny, and R. A. Schmidt. Computerized detection of masses in digital mammograms: automated alignment of breast images and its effect on bilateral-subtraction technique. *Med Phys*, 21:445–452, Mar 1994.

[177] T. S. Yoo. *Insight into Images: Principles and Practice for Segmentation, Registration, and Image Analysis*. AK Peters Ltd, 2004.

[178] P. A. Yushkevich, J. Piven, H. C. C. Hazlett, R. G. G. Smith, S. Ho, J. C. Gee, and G. Gerig. User-guided 3d active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *NeuroImage*, 31(3):1116–1128, July 2006.

[179] L. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338 – 353, 1965.

[180] B. Zangheri, C. Messa, M. Picchio, L. Gianolli, C. Landoni, and F. Fazio. PET/CT and breast cancer. *European Journal of Nuclear Medicine and Molecular Imaging*, 31:S135–S142, 2004.

[181] Y. Zheng, S. Englander, S. Baloch, E. I. Zacharaki, Y. Fan, M. D. Schnall, and D. Shen. STEP: spatiotemporal enhancement pattern for MR-based breast tumor diagnosis. *Med Phys*, 36:3192–3204, Jul 2009.

[182] A. A. Zytoon, K. Murakami, M. R. El-Kholy, E. El-Shorbagy, and O. Ebied. Breast cancer with low FDG uptake: Characterization by means of dual-time point FDG-PET/CT. *European Journal of Radiology*, 70(3):530 – 538, 2009.