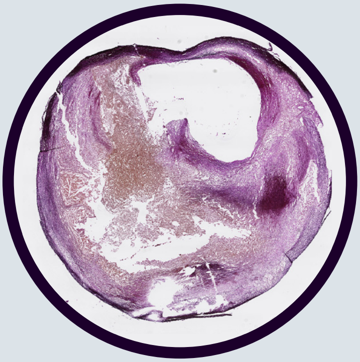
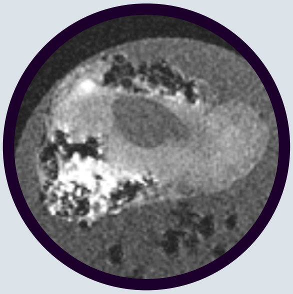
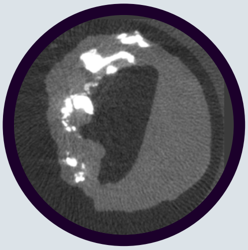
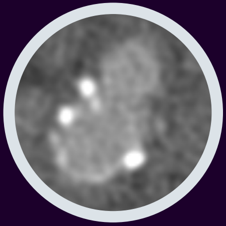
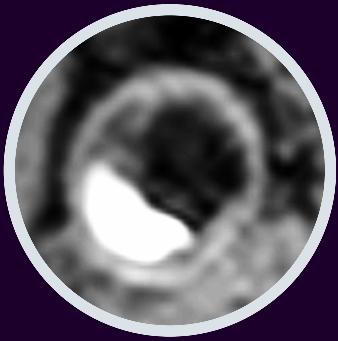


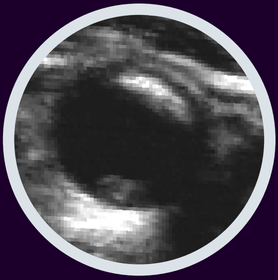




Multimodal Image Analysis for Carotid Artery Plaque Characterization







Arna van Engelen

Multimodal Image Analysis for Carotid Artery Plaque Characterization

Arna van Engelen

The work in this thesis was conducted at the Departments of Radiology and Medical Informatics of the Erasmus MC, University Medical Center, Rotterdam, the Netherlands.

The research described in this thesis was performed within the framework of CTMM, the Center for Translational Molecular Medicine (www.ctmm.nl), project PARISk (grant 01C-202), and supported by the Dutch Heart Foundation (DHF-2008T094).

Advanced School for Computing and Imaging

This work was carried out in the ASCI and COEUR graduate schools.
ASCI dissertation series number 304.

Financial support by the Dutch Heart Foundation for the publication of this thesis is gratefully acknowledged.

Additional financial support for printing of this thesis was kindly provided by the department of Radiology, Erasmus MC; ASCI; and Medis medical imaging systems bv.

FSC
www.fsc.org
MIX
Papier van
verantwoorde herkomst
FSC® C021795

Printed by GVO drukkers & vormgevers B.V. | Ponsen & Looijen

ISBN: 978-90-6464-774-1

Cover: examples of the different images that have been used in this thesis. At the same time it represents a linear classifier, in which the samples have different weights indicated by their size.

Multimodal Image Analysis for Carotid Artery Plaque Characterization

Multimodale beeldanalyse voor halsslagaderplaque-karakterisatie

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

prof.dr. H.A.P. Pols

en volgens het besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
dinsdag 17 juni 2014 om 11.30 uur

door

Arna van Engelen

geboren te Heesch

ERASMUS UNIVERSITEIT ROTTERDAM

Promotiecommissie

Promotor:	Prof. dr. W.J. Niessen
Overige leden:	Prof. dr. ir. A.F.W. van der Steen Prof. dr. B. van Ginneken Prof. dr. A. Fenster
Copromotor:	Dr. M. de Bruijne

Contents

Contents

Chapter 1

Introduction

Cardiovascular diseases are widespread, and the leading cause of death worldwide [9]. Moreover, they largely contribute to worldwide morbidity and disability, and thereby healthcare costs [180, 235]. Cardiac and cerebrovascular ischemia are two of the largest subtypes of cardiovascular disease. These ischemic events often result from the presence of atherosclerosis [128], also known as hardening of the arteries. Unfortunately, atherosclerosis is difficult to detect in an early stage: about half of the cases of sudden cardiac death occurs without known presence of cardiac heart disease [66]. In order to prevent vascular events from happening, accurate risk stratification of patients at increased risk has great value and enables targeting treatment to the right patients.

Atherosclerosis affects the vessel wall, causing wall thickening and lumen narrowing. These can both be assessed by imaging. Especially the carotid artery is suitable for non-invasive imaging by several modalities, as it is closer to the skin than the aorta and larger than the coronary arteries. In this thesis we focus on automated image analysis for characterization of the atherosclerotic carotid vessel wall, to help predict the risk of ischemic events. This introduction first gives a background on the carotid artery and atherosclerosis (Section 1.1), vessel wall imaging (Section 1.2), and image processing for vascular disease (Section 1.3), followed by an outline of the remainder of this thesis (Section 1.4).

1.1 Carotid atherosclerosis

The carotid artery is the artery that supplies the head with blood and oxygen (Figure 1.1a). The common carotid artery branches off from the aorta towards the head. At the bifurcation the vessel splits into the external carotid artery which supplies the face, and the internal carotid artery, which supplies the brain with oxygen-rich blood.

Atherosclerosis is related to aging of the vascular system and affects all large arteries of the body. Although it is usually diagnosed in one vessel first, it is a systemic disease and disease in different body parts is related. It has been shown that examining the carotid artery can give an indication of the general state of atherosclerotic disease in the rest of vascular system [90, 185].

1.1.1 Development of atherosclerosis

Atherosclerosis is mostly detected in individuals over 45 years of age, but studies have shown that development already starts at a much earlier point in life [65, 127, 212]. The slow process of atherosclerosis development is shown in Figure 1.1b. It starts with outward thickening of the vessel wall, and is in a later stage followed by inward thickening as well, which causes luminal stenosis. This process is generally assumed to be initiated by inflammatory processes of the endothelial vessel wall cells in response to cholesterol molecules. This starts a chain reaction in which more and more cholesterol and macrophages accumulate in the vessel wall and the muscle cells start to enlarge. In a later stage calcification develops, and the accumulated cholesterol molecules form a lipid-rich, often necrotic core, that is covered by a fibrous cap.

Risk factors for the development of atherosclerosis are among others increased age, blood pressure, cholesterol level, body mass index, and smoking. The specific location at which a plaque develops furthermore relates to the vascular geometry [59, 174, 177]. The inner curvature of vessels and bifurcations, among which the carotid bifurcation, are predominant locations. A possible causal mechanism for this is a low mechanical wall shear stress, which was shown to alter cell properties with a subsequent increase in lipid uptake in those regions [48, 174].

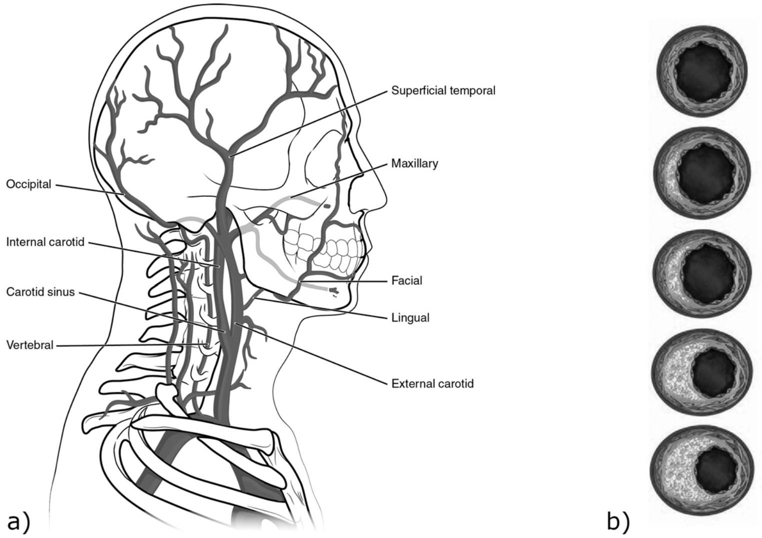


Figure 1.1: a) Anatomy of cerebral vessels including the carotid arteries (image from Connexions, *http://cnx.org/content/m46646/latest/)*. b) Schematic example of plaque development from top to bottom, starting with outward remodelling followed by inward remodelling causing stenosis (image from AstraZeneca).

1.1.2 Ischemic risk

In this thesis we focus on atherosclerosis of the carotid artery. More than in complete occlusion of this artery, the danger of atherosclerosis lies in the possibility of plaque rupture. At the moment of rupture plaque tissue breaks loose from the vessel wall and enters the blood stream. In case of the carotid artery this may cause embolization of thrombus or the plaque material in the brain and subsequent cerebral ischemia. For a long time the risk of such events to occur has been related to the degree of stenosis and wall thickening. However, patients with similar stenosis degree and wall thickening were found to have different risks.

Histological study of excised plaques from symptomatic versus asymptomatic patients has shown that plaque composition and structure relate to plaque vulnerability and have led to the concept of the 'vulnerable plaque' [55]. Ruptured carotid plaques have larger lipid cores, thinner fibrous caps, more macrophage infiltration, more intraplaque hemorrhage and less calcification [167, 168]. These features as detected by non-invasive imaging have also been shown to increase the risk for ischemic events in longitudinal studies [70, 160, 188].

Biomechanical parameters have also been studied in the context of the relation between plaque composition and plaque rupture [67, 175]. Rupture-prone plaques occur more often upstream of the stenosed region, where wall shear stress is higher. The hypothesis is that eventual rupture is not caused not by wall shear stress, but occurs when peak plaque cap stress induced by arterial pressure exceeds the fibrous cap's strength. Large stresses at the cap can develop, depending on plaque morphology and biomechanical properties of the tissue components: peak cap stress increases with a thinner cap, a larger necrotic core and smaller necrotic core angle [6]. Moreover, macrophage infiltration weakens the fibrous cap [114].

1.1.3 Treatment

Different treatment options for carotid atherosclerosis are available. In an early stage lifestyle modifications (cessation of smoking, changing diet, changing physical activity) are recommended, but also medication (anticoagulants, antihypertensive drugs or lipid-lowering drugs including statins), and in a later stage surgical intervention are possible. The recommendation for surgical intervention is currently based on the degree of carotid artery stensosis as determined by non-invasive imaging [26, 143, 152, 197]. Imaging technologies used are duplex ultrasound, X-ray angiography, CT-angiography or MR-angiography, or a combination of these. Intervention is generally recommended for symptomatic patients with a stenosis degree of 70-99% of the carotid artery. In case of asymptomatic stenoses >70% some guidelines recommend intervention as well, but its advantage over best medical treatment available is doubtful [26, 143]. Intervention can be either carotid endarterectomy (CEA), which is the excision of plaque from the vessel, or carotid artery stenting (CAS) to widen the vessel. CEA is more commonly used. It has been associated with a lower risk of complications than CAS in symptomatic patients. In asymptomatic patients no difference was seen, however, CAS is more expensive [144].

Although at this point the guidelines for the management of patients with carotid atherosclerotic disease only take the degree of stenosis into account, some indicate that in the future inclusion of more characteristics will be important. Brott et al. [26] mention that plaque composition may be more important than the degree of stenosis, and that the search for other predictors must advance, and Ricotta et al. [152] state that plaque morphology is an important feature to assess future risk, but that it 'requires further prospective evaluation to determine its ultimate clinical utility'.

A few studies have already been evaluated in which adjustment of lifestyle and medication was guided by vessel wall imaging as opposed to traditional risk factors and lumen imaging. Intensification of medical treatment of patients with progression of plaque area on 2D ultrasound was shown to decrease the average rate of plaque progression, and the amount of cardiovascular events among asymptomatic patients with stenosis >60% [181, 183]. Moreover, treatment selection (CEA or CAS) based on presence of a high-intensity signal on preoperative Time-of-Flight MR-angiography, as a marker of intraplaque hemorrhage, significantly reduced the amount of periprocedural events [242].

1.2 Vessel wall imaging

Imaging can be used to to detect atherosclerosis, to monitor disease progression, and to determine whether intervention is warranted. It can be performed to assess the geometry and stenosis degree of the vessel lumen, but also to assess vessel wall thickening, and the composition of the diseased vessel wall. Whereas imaging is already used in clinical practice to assess the degree of luminal stenosis [120, 143], and to some extent wall or plaque thickness [183], we focus here on imaging to assess wall composition. Both invasive and non-invasive modalities for vessel wall imaging have been developed. Promising invasive imaging modalities that have mainly been used to study the coronary arteries include intravascular ultrasound [137, 201], optical coherence tomography [194] and photo-acoustic imaging [91, 92]. Here we focus on and discuss the three main non-invasive imaging modalities that are used to study the carotid artery wall: Ultrasound, Computed Tomography (CT) and CT-angiography (CTA), and Magnetic Resonance Imaging (MRI). Naturally, each modality has its advantages and disadvantages [107, 161, 166]. Examples of images of the carotid artery visualized by these three modalities are shown in Figure 1.2.

1.2.1 Ultrasound

The carotid artery lies close to the skin and is therefore an ideal structure to image with ultrasound. Due to the relatively low costs, wide availability and lack of ionizing radiation ultrasound is especially suitable in screening for atherosclerosis [165]. To acquire an ultrasound scan a transducer is placed on the skin. The transducer emits pulses of high-frequency sound waves (>20 kHz). These waves interact with the tissue: reflections occur mainly at tissue boundaries due to density changes, and acoustic scattering occurs depending on tissue properties. Echoes from wave interactions with tissues are detected. In traditional B-mode imaging the detected echoes from multiple scan-lines are combined into 2D images. A trade-off between resolution and penetration depth determines the ultrasound frequency used. With higher frequencies and shorter wavelengths, smaller pixels are obtained, but with limited penetration depth. For carotid ultrasound an axial resolution with pixel sizes of 0.1-0.3 mm is common.

In addition to traditional 2D B-mode ultrasound imaging, 3D ultrasound acquisition is becoming more popular [52, 53]. Either by electromagnetic tracking of the device in a free-hand acquisition, or by using a mechanical transducer, acquired 2D slices can be combined into a 3D volume. This allows for the interpretation of the full vessel without being dependent on the acquired 2D slice. The two ultrasound images in Figure 1.2 are selected sections from a 3D volume.

Traditionally vessel wall thickness measured as intima-media thickness (IMT) from 2D ultrasound has been used as a measure of atherosclerotic disease [121], and IMT change has been used as a measure of disease progression. However, recent meta-analyses showed that IMT can only marginally improve risk stratification over common risk factors in the Framingham risk score [47] and that IMT change does not predict vascular events [122]. A shift has been made towards measuring changes in plaque area or 3D plaque volume, which were shown to be stronger predictors [229]. Moreover, the appearance of the vessel

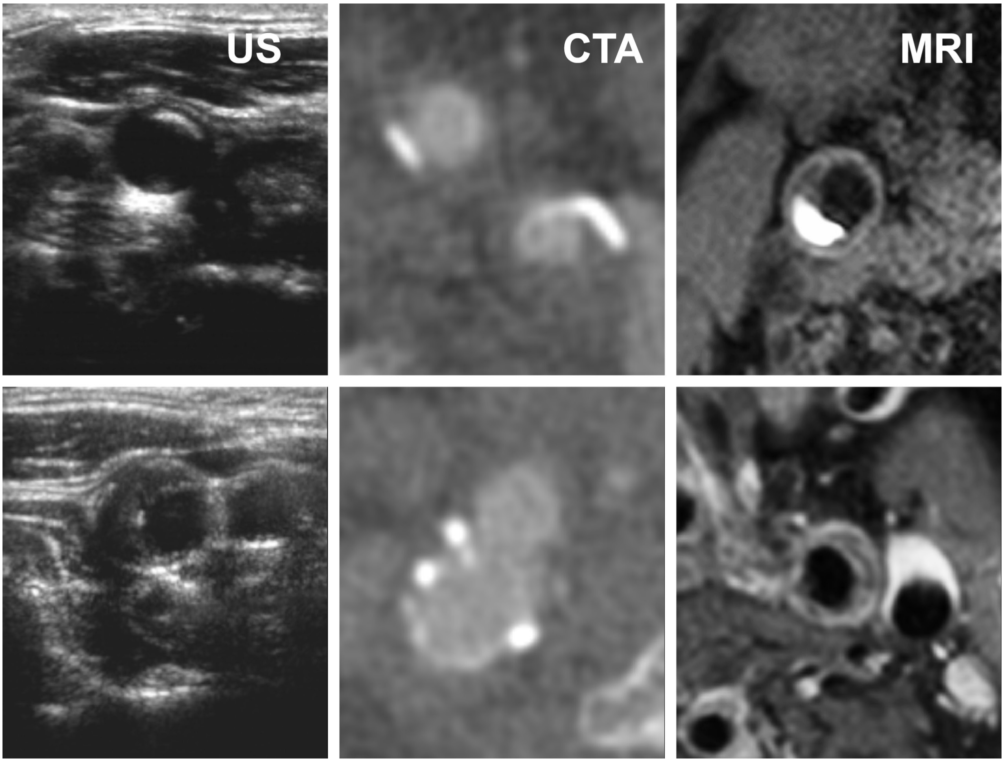


Figure 1.2: Examples of axial slices of the carotid artery in ultrasound (US), CT-angiography and MRI. The ultrasound images show a plaque with high echogenicity in the top row and low echogenicity in the bottom row. CTA clearly depicts the lumen and calcification. The T1 IR-TFE MRI in the top row has a high-intensity region of intraplaque hemorrhage and a dark calcification spot. The postcontrast T1w MRI in the bottom row shows a hypointense region in the center of the thickened wall due to a lipid-rich necrotic core.

wall, defined by plaque texture, has been related to plaque composition. Fibrous tissue is more echogenic and thus has higher image intensities than regions of lipid and hemorrhage which are more echolucent [51, 110]. Calcifications cause strong echoes and thereby may cause acoustic shadowing which limits analysis of the plaque lying behind the calcification. Figure 1.2 shows an echogenic plaque in the top row and an echolucent plaque in the bottom row.

1.2.2 CT

CT is an imaging modality that creates 3D image volumes using tomographic reconstruction of x-rays [87]. Image intensity is measured in standardized Hounsfield units and results from x-ray attenuation by the tissue. CTA uses injection of a contrast agent and is used to image the vasculature. Current systems are able to perform rapid imaging, and a full CTA image of the carotid arteries with a voxel size of $\sim 0.3 \times 0.3 \times 0.5$ mm can be acquired in the order of seconds [44]. A disadvantage of CT is the use of ionizing radiation, which is assumed to impose long-term risks on scanned subjects [25]. Therefore, current improvements in CT imaging and image analysis are often aimed at dose-reduction [95, 151].

CTA is commonly used for carotid artery imaging [84, 215]. It can be used for accurate stenosis grading and for quantification of arterial wall calcification. Calcification has high intensity in CT, as the examples in Figure 1.2 show. The adequacy of CT for depicting other plaque components has been studied as well [45, 237]. Some studies found a difference in Hounsfield units between fibrous and lipid or necrotic tissue [45]. However, the accuracy with which those components can be distinguished in CTA remains questionable [237].

1.2.3 MRI

MRI [112] has better soft tissue contrast than ultrasound and CT, and is therefore more suitable for imaging different plaque components. In contrast to CT, no ionizing radiation is used. Instead, the object or person to be scanned is placed in a strong magnetic field (1.5-3 Tesla in general clinical practice). Images are created by briefly applying a different electromagnetic field to a region of interest. The behaviour in response to this pulse is tissue-dependent, and therefore tissues with different magnetization properties can be distinguished. By measuring at different time intervals after the electromagnetic pulse different contrasts are generated, which is useful in many applications and essential to image all plaque components of interest.

The contrast and appearance of different plaque components in MRI has been studied well [159, 202, 243]. By comparing to histology it has been determined that calcification has low signal intensity in all MRI sequences, and that intraplaque hemorrhage appears bright in T1-weighted images. Lipid-rich necrotic tissue can be visualized in several ways. The best contrast is obtained when pre-contrast and post-contrast T1-weighted images are compared [187, 246]. Contrast agent is taken up by fibrous tissue but not by necrotic tissue, resulting in a hypointense region post-contrast which is iso-intense precontrast at the location of a necrotic core. When no contrast agent is used lipid or necrotic tissue can be recognized by hypointensity in T2-weighted or diffusion weighted MRI [38]. Several approaches of imaging plaque inflammation by means of nanoparticles have been studied as well [96, 193]. Examples of different MR contrasts are shown in Figure 1.2.

1.3 Image processing

Manual analysis of images is time-consuming and labour-intensive. In addition, there can be considerable variation between observers. Especially when quantitative measurements have to be performed, human workload is a limiting factor for usage in both large-scale research projects and in clinical practice. Automated image analysis is therefore crucial in these situations.

Automated segmentation of the carotid artery lumen and outer wall has been the topic of studies in ultrasound [13, 208, 241], CTA [73, 126, 227] and MRI [14, 72, 209, 218]. Most of these methods obtain reasonable segmentations, although manual adjustments especially in the bifurcation and plaque regions remain necessary to obtain segmentations that can be used for accurate grading of stenosis, wall thickness or further processing [73]. Assuming a correct wall segmentation, characterization of the vessel wall content is the next step and the focus of this thesis.

In ultrasound most studies aiming to characterize the vessel wall have focused on gray scale median (GSM) of the vessel wall. GSM is moderately correlated with histology composition [236], and was shown to be predictive of vascular events [69, 149]. More advanced texture measures have been studied as well. A combination of local binary pattern and Law's features [2], and a set of 56 texture features [35] were shown to enable classification of symptomatic versus asymptomatic subjects, and Haralick texture features predicted future cerebrovascular events by support vector machine classification [109]. We study the predictive value of a large set of texture features in this thesis. An attempt to segment plaque components based on gray scale has been made as well [110]. Although this approach has not been followed on B-mode ultrasound, a similar approach has been used to develop a so-called virtual histology method for intravascular ultrasound, which is more commonly used in research [134].

For CTA thresholds based on Hounsfield values have been determined after manual alignment of CTA with histology slices, to separate lipid, calcium and fibrous tissue [45]. This yielded good correlations between histology and CTA segmentations of calcium and fibrous volumes, but a poor correlation for lipid especially in more calcified plaques. A different study used a similar approach and found perfect agreement for calcification, but due to large overlap in intensity between lipid-rich necrotic core, fibrous tissue, and hemorrhage there was limited agreement for these components [237]. Still, significant correlations between the automatic segmentation results as determined by [45] and clinical parameters were found [226].

Due to its superior soft tissue contrast, most work on segmentation of plaque components has been performed in MRI. MRI is more challenging than CT in the sense that there is no standardized intensity such as the Hounsfield scale, and even within images intensity inhomogeneity occurs. In addition, owing to the complementary information available in different MR images to assess all relevant plaque components it is important to combine different MRI contrast images. Therefore, image registration, image normalization and inhomogeneity correction are needed to obtain good results in a subsequent segmentation framework. In *ex vivo* MRI good segmentation results with respect to histology have been obtained based on voxel classification of multicontrast MRI intensities [37, 38, 89, 97, 154]. In *in vivo* MRI reasonable to good results have been obtained as well [82, 119, 219]. While Liu et al. [119] obtained good accuracy with respect to histology for all studied components (calcium, necrotic core, loose matrix and fibrous tissue), van 't Klooster et al. [219] and Hofman et al. [82] obtained good results for most components, but only poor to moderate agreement with the ground truth for calcification. These methods use a voxel classification approach with a Bayesian or linear discriminant classifier, where in addition to intensities, spatial information [119, 219] and multiscale and multiorder Gaussian derivatives [219] were included.

For segmentation methods based on voxel classification supervised methods are preferred, because they generally can yield a better performance. These require an accurate ground truth. For plaque segmentation several approaches have been used to obtain such ground truth. Histology sections contain the most accurate information on tissue composition, but these are not always available and difficult to register with *ex vivo*, but especially *in vivo* imaging data. Therefore also manual annotations and histology-guided manual annotations have been used.

A limitation of most current supervised segmentation methods is that while they may perform very well on data with the exact same properties as the data used for training, accuracy quickly decreases when the data to analyze is different in for example scanner type, imaging protocol or patient population. For MR brain segmentation, a number of methods have been proposed and evaluated to overcome this problem [39, 56, 216, 217]. In this thesis we present a transfer-learning approach similar to Wu et al. [238] and Van Opbroek et al. [216] for the carotid artery wall.

Besides single-modality analysis, several studies have showed an interest in combining MRI with either CT [74, 132] or ultrasound [34] for a better description of the vessel wall. These studies suggest that a combination of imaging modalities may better determine whether a plaque is vulnerable, or that one imaging modality could be used to determine which patients benefit from subsequent imaging by a second imaging modality. We evaluate different imaging modalities and combine MRI and CT in this thesis.

1.4 Outline of this thesis

Given the described previous work and the available literature, it is relevant to study characterization of the atherosclerotic vessel wall using supervised methods, and to study the use of multimodal imaging. Therefore this thesis describes the development and validation of techniques that aim at the automated characterization of the carotid atherosclerotic vessel wall. Both methods for *ex vivo* and *in vivo* imaging modalities are developed, including MRI, CTA, ultrasound and histology.

Chapters 2 to 4 make use of a dataset in which *ex vivo* and *in vivo* MRI and CT, and manually annotated histology sections are available, that have largely been registered. In **Chapter 2**, we perform voxelwise classification of calcification, lipid-rich necrotic and fibrous tissue in the *ex vivo* MRI, using a ground truth obtained by registering histology and μCT images with the MRI. We investigate the discriminative value between classes when different groups of features are included: intensities, Gaussian filters and lumen and outer wall distances. In addition, to evaluate the advantage of using an extensive framework for 3D registration instead of previous methods, we compare segmentation accuracy between the situation where all imaging modalities are registered using 3D registration, and slice-based matching of selected histology slices with *ex vivo* MRI slices.

In **Chapter 3** we develop voxelwise classification algorithms for the combination of *in vivo* MRI and CTA. This chapter handles the problems that arise due to difficult alignment of histology with *in vivo* data. Firstly, some additions to the image registration framework are made. Secondly, to account for errors in ground truth labels by misregistration we evaluate three different ways of sample weighting: 1) a weighting of samples per slice based on registration accuracy per slice, 2) a weighting that weighs each sample individually based on local registration accuracy, and 3) outlier rejection. In addition, we compare segmentation accuracy based on the combination of MRI and CTA to use their complementary advantages, to the use of only CTA and only MRI.

For the methodology in both Chapters 2 and 3 an accurate histology segmentation is required. Manual annotation is labor-intensive and time-consuming. **Chapter 4** proposes a method for (semi-)automatic segmentation of histology slices. We compare training on the resulting automatically segmented histology slices for both *ex vivo* and *in vivo* image segmentation (as in Chapters 2 and 3) to training on the manually segmented histology. As histological image appearance varies greatly between specimens, fully automatic segmentation by training on independent data is compared to an interactive approach with training on 1 or 2 slices of the histology specimen to be segmented.

In **Chapter 5** we develop an algorithm to segment plaque components in *in vivo* MRI in the multi-center *PARISk* study [205]. Data with manual annotations from two centers is used for this chapter. In both protocols five image sequences are acquired to distinguish plaque components, but scanner vendor, coil hardware and protocol implementation differ between the centers. We compare the accuracy of plaque component segmentation between the two protocols, and evaluate to what extent an algorithm trained on one protocol can be applied to segment data acquired in the other center. Methodologies to improve this transferability, including *transfer-learning* methods, are presented to improve the performance of classification algorithms in this context.

In **Chapter 6** we aim to use texture analysis of plaques in 3D ultrasound to predict vascular events. Instead of performing a segmentation of individual plaque components, we calculate overall plaque texture per patient. In this longitudinal study patients were scanned at baseline and after one year, and followed for vascular events for up to five years. A large number of texture features is acquired for all plaques at both time points. Sparse Cox regression is used to reduce dimensionality and to predict plaque vulnerability per patient. The predictive value of texture, change in texture and change in plaque volume, and combinations of those measures, is evaluated.

This thesis is concluded with a summary of all findings, a discussion of these findings and a future perspective in **Chapter 7**.

Chapter 2

Multi-feature-based plaque characterization in *ex vivo* MRI trained by registration to 3D histology

This chapter has been published as:
A. van Engelen, W.J. Niessen, S. Klein, H.C. Groen, H.J.M. Verhagen, J.J. Wentzel, A van der Lugt and M. de Bruijne, Multi-feature-based plaque characterization in *ex vivo* MRI trained by registration to 3D histology, *Physics in Medicine and Biology*, 2012

* * *

Abstract
We present a new method for automated characterization of atherosclerotic plaque composition in *ex vivo* MRI. It uses MRI intensities as well as four other types of features: smoothed, gradient magnitude and Laplacian images at several scales, and the distances to the lumen and outer vessel wall. The ground truth for fibrous, necrotic and calcified tissue was provided by histology and μCT in twelve carotid plaque specimens. Semi-automatic registration of a 3D stack of histological slices and μCT images to MRI allowed for 3D-rotations and in-plane deformations of histology. By basing voxelwise classification on different combinations of features, we evaluated their relative importance. To establish whether training by 3D registration yields different results than training by 2D registration, we determined plaque composition using 1) a 2D slice-based registration approach for three manually selected MRI and histology slices per specimen, and 2) an approach that uses only the three corresponding MRI slices from the 3D-registered volumes. Voxelwise classification accuracy was best when all features were used (73.3±6.3%), and was significantly better than when only original intensities and distance features were used (Friedman, $p<0.05$). Although 2D registration or selection of three slices from the 3D set slightly decreased accuracy, these differences were non-significant.

2.1 Introduction

Cerebral infarction is one of the most important causes of death and the largest cause of disability in the Western world. Approximately 20% to 30% of all cerebral infarctions can be related to atherosclerotic plaque rupture in the carotid artery [146]. Evidence suggests that plaques prone to rupture are characterized more by plaque composition than by the plaque size or the degree of stenosis [167, 168]. Magnetic resonance imaging (MRI) is able to non-invasive visualize different plaque components. Both *in vivo* [36, 77, 129, 158, 202, 243] and *ex vivo* [130, 170] the accuracy of MRI to distinguish plaque components has been demonstrated by comparison with histological sections. Relating plaque composition to plaque rupture or clinical events showed that a large lipid or necrotic core [167] covered by a thin fibrous cap [244] is associated with a high risk of rupture, while fibrous tissue and calcifications [168] have a stabilizing effect. In addition, the presence of intraplaque hemorrhage [188] and plaque ulceration [145] are related to an increased number of cerebrovascular events. Most studies analyzed the relation between the presence of plaque components and outcome. However, the relative size of the components is also important for risk prediction. Segmentation and quantification of plaque components in MR images is therefore relevant. In addition, segmentations can be used to make biomechanical models of plaques [6]. Automatic methods may reduce the inter- and intraobserver variability and are more practical in large studies. Our study focuses on the automatic segmentation of plaque components in *ex vivo* MRI.

Automated plaque-segmentation methods have previously been presented by several authors, both in *ex vivo* [37, 38, 89, 97, 154] and *in vivo* MRI [3, 82, 119]. Most of these methods are based on voxelwise statistical classification. For *ex vivo* MR images of plaque specimens, segmentations of different components have been obtained with a maximum-likelihood classifier [37], a nearest-mean classifier [38], k-means clustering [97], and k-means clustering in combination with an iterative approach to minimize pixel discontiguity [89]. The probability of misclassification was similar for a normal density based approach and non-parametric Parzen window estimation [154]. Similar approaches have been introduced for the analysis of *in vivo* MR images. In a comparison, a standard Bayesian classifier performed better than k-nearest neighbour, a feedforward neural network, and a Bayesian classifier in combination with a Parzen classifier [82]. In addition, fuzzy clustering followed by contour smoothing [3], and a maximum-likelihood Bayesian classifier that besides intensities also included the distance of each voxel to the lumen and the local wall thickness [119] have been investigated. Except for this last method [119], all these methods are based only on original MRI voxel intensities. This paper investigates whether additional image features contain information that improves classification accuracy, validated by a histological ground truth.

To evaluate plaque segmentation, the resulting areas or volumes can be compared to those resulting from manual delineations in the MRI [3], to volumes obtained from histology [82, 97], or to the visual grading of the American Heart Association (AHA) [89]. A disadvantage is that these approaches use global measures and do not take account of regional correspondence between the classification result and the ground truth. Other studies used co-registration between histology and MRI to perform voxelwise analysis [37, 38, 119, 154]. Corresponding histology and MRI slices were then manually selected prior to in-plane registration, and voxelwise classification was evaluated for a selection of on average 1 to 5 slices per specimen. This manual selection may positively bias classification accuracy as the matched MRI and histology slices may appear more similar and have higher image quality.

On the other hand, with 2D in-plane registration it is not possible to account for differences in out-of-plane slice orientation between the histological and MRI slices. Therefore, we aim to develop a classification method that is based on registration of MRI to 3D histology [68]. We expect this to provide a more accurate and objective ground truth, and it allows the inclusion of MRI slices with a lower image quality in both training and evaluation.

This study presents a new method for segmenting calcification, fibrous and lipid-rich necrotic tissue in MR images obtained from *ex vivo* carotid artery specimens. As well as using MRI intensity and distances to the lumen and outer vessel wall, we evaluated Gaussian smoothed intensities, gradient magnitude and Laplacian images for their contribution to classification accuracy. A 3D registration of MRI with histology was used for classifier training and evaluation. This approach was compared to the more standard method of 2D in-plane registration of selected MRI and histology slices. Preliminary results of this study have previously been published in a conference proceedings [213].

2.2 Materials and methods

2.2.1 Data

Fifteen patients scheduled for carotid endarterectomy (CEA) were included in the study (9 had an ischemic stroke, 5 had a transient ischemic accident, and 1 was asymptomatic). During the CEA procedure, the plaque was removed intact as much as possible to keep its 3D shape. The *ex vivo* specimens were scanned in a 3T clinical MRI scanner (Signa Excite, GE Healthcare, Milwaukee, USA). Three MRI contrast weightings were obtained: 3D-T1 weighted (T1w) gradient echo (GRE), 2D-T2 weighted (T2w) spin echo (SE) and 2D-Proton-density weighted (PDw) SE. Scan parameters are provided in Table 2.1. To allow accurate calcification quantification, the specimens were also scanned with a μCT scanner (Skyscan 1072, Skyscan, Belgium, resolution $18\times18\times18\,\mu$m).After imaging, the specimens were decalcified and embedded in paraffin for histological processing. At every 1mm interval, axial slices of 5 μm thickness were obtained. In addition, digital photographs of the cutting plane, called 'enface', were taken to allow reconstruction of the 3D histology volume. An Elastica von Gieson stain was applied (EvG, Merck, Germany), to visualize cell nuclei, elastin and collagen fibers. On the high-resolution digitized histologic slices, contours of the lumen and outer vessel wall and for regions of lipid-rich necrotic tissue were drawn manually. Calcified areas were obtained from the μCT by intensity thresholding. The μCT scanner was not calibrated and values could not be related to standard Hounsfield units. Therefore one threshold value was chosen that gave visually satisfying results for all specimens. As the resolution and contrast for calcification are both quite high in μCT, the exact threshold value is not very critical. The remaining areas were considered to be fibrous tissue, resulting in a ground truth with three components (calcification, fibrous tissue and lipid-rich necrotic tissue). Three specimens were excluded for classification due to a low quality of histology (2 cases) or an incomplete MRI protocol (1 case), leaving 12 specimens. We excluded slices for which it was not possible to obtain a ground truth due to histology disruption. The number of histology slices with corresponding MRI included was 11±4 per specimen (range 5 - 17). Figure 2.1 shows a flow chart of the methods in this paper. It includes an example for the ground truth obtained in one image slice.

Table 2.1: MRI scan parameters.

Sequence	Resolution (mm)	Repetition time (ms)	Echo time (ms)	Flip angle (degrees)
3D-T1w GRE	0.1x0.1x0.1	32.6 - 57.1	3.6 - 4.9	60
2D-T2w SE	0.1x0.1x0.6	2300 - 13820	53.9 - 67.0	90
2D-PDw SE	0.1x0.1x0.6	7700 - 14860	17.9 - 18.3	90

2.2.2 Registration of MRI with histology

3D registration

The 3D registration procedure followed a pipeline that was developed to match *in vivo* CTA with histology, via μCT and *ex vivo* MRI, using the same settings as published previously [68]. This procedure includes multiple steps using both rigid and deformable registration, and involves several manual annotations of anatomical landmarks, lumen and outer vessel-wall contours in each image modality. The top row in Figure 2.1 illustrates the different stages in registration. Here, a more detailed description of each step is given.

- *Enface stacking*: The 'enface' images taken during histology slicing were stacked by 2D in-plane point-based rigid registration of manually annotated landmarks to form

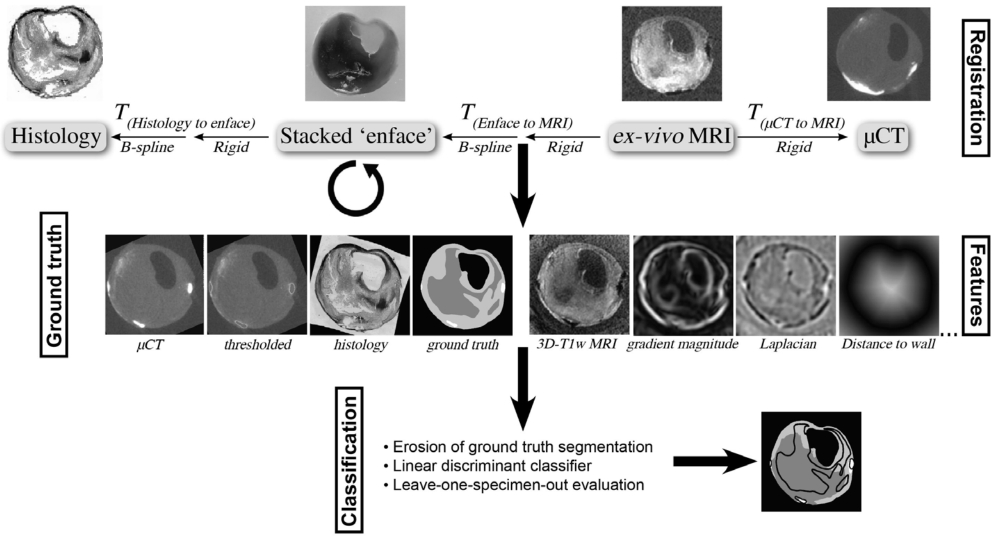


Figure 2.1: Flow chart of the method. Registration: Images before registration are shown, the arrows indicate what registration steps are applied and point from the fixed to the moving images. The large arrow points down from the domain in which all images are registered with each other and classification is done. Here μCT and MRI are only rigidly deformed, and the histology is non-rigidly deformed to match the *ex vivo* MRI. The in-plane resolution equals the 'enface' resolution, which is 0.017×0.017 mm. Ground truth: deformed μCT (original and thresholded) and histology are shown with the resulting ground truth. Contours are shown in the undeformed histology slice in the upper row. Features: a slice of the transformed 3D-T1w MRI with its gradient magnitude and Laplacian image and the distance to the outer vessel wall for each voxel (black=0) are shown. Classification: a segmentation result is shown (white=calcification, light grey=fibrous tissue, dark grey=necrotic tissue.

a 3D volume. Each 2D image was registered to its preceding image (8±3 landmarks per 2D registration were used).

- $T_{(Histology\,to\,enface)}$: The 2D histology slices were registered to the enface stack to obtain a 3D histology stack, allowing for non-rigid in-slice deformations. Manually annotated landmarks (5±1 per slice) were used for rigid initialization, and the B-spline model [157] was used to maximize the mutual information (MI) [199, 222] between the original images, manual lumen segmentations, and manual outer-wall segmentations in both the 'enface' and histology images.
- $T_{(Enface\,to\,MRI)}$: This registration was initialized with a rigid registration based on manually annotated landmarks in the 'enface' and 3D-T1w MRI (8±2 landmarks per specimen). A rigid registration of the MRI with the 3D histology stack was then performed with the same MI maximization and similarity measure as before, for which manual lumen and outer vessel wall segmentations were made on the MRI. The registration was refined by B-spline deformable registration with the same similarity measure, in which only in-plane deformations were allowed while maintaining continuity along the slice direction. Since the specimen did not move between acquisitions, the different MR images were assumed to be aligned with each other.
- $T_{(\mu CT\,to\,MRI)}$: To register the μCT to the *ex vivo* MRI, a point-based rigid registration without scaling was performed using annotated landmarks in the μCT and 3D-T1w image (7±2 landmarks per specimen).

All registration steps were performed using the registration toolbox `elastix` [102]. Classification was performed in the space inbetween the *ex vivo* MRI and stacked histology, in Figure 2.1 indicated by the large arrow pointing down from the registration part. All images were transformed to this domain. This means that only the B-spline transformation of $T_{(Enface\,to\,MRI)}$ was applied to the 3D histology stack, and that the inverse rigid transformation was applied to the *ex vivo* MRI. The in-plane resolution in this domain was chosen equal to the resolution of the enface images, 0.017×0.017 mm, with a slice distance of 1 mm. In this way, interpolation between histology slices was avoided.

2D registration

To study whether the match between MRI and histology with either 3D or 2D registration affects the classification results, two additional datasets were created. For that purpose, an observer manually matched 2D image slices from the MRI volumes to all histology slices of each specimen, using a procedure similar to the method used in several papers [37, 38, 154]. Matching was achieved using the known distance between slices in both MRI and histology, and the shape of the lumen and outer vessel wall. The selected MRI slices had the same position in each of the three contrast weightings, as the specimen did not move between acquisitions. As the studies cited had used fewer slices per plaque for classification (1.25, 3 and 3-5), for this study the observer selected three slices from each specimen. The slices that had the best histology and MRI quality and covered as much as possible of the length of the specimen were selected, excluding adjacent slices. The selected MRI slices were registered in-plane with the corresponding histology, as in $T_{(Enface\,to\,MRI)}$ for 3D registration: using a point-based rigid initialization (6±2 landmarks per slice) followed by rigid and B-spline deformations that optimized the MI between the original images, manual lumen segmentations, and manual wall segmentations in both images.

To better compare classification results in these 3D- and 2D-registered datasets, a third dataset was created. This consisted of the three MRI slices out of the 3D-registered dataset that corresponded to the three histology slices that had been selected for the 2D-registered dataset.

2.2.3 Voxel classification method

To remove intensity nonuniformities in the original MR images, we used nonparametric nonuniform intensity normalization (N3) [176]. For intensity normalization between patients, the intensity values were scaled with respect to the average intensity of the fluid in the tube surrounding the plaque, which was set to 1000. To account for slight misregistrations and partial-volume effects, both in classifier training and evaluation a morphological erosion with a disk-structuring element was applied to the segmentations. For calcifications, the diameter of the structuring element was set to 3 voxels (0.05 mm); for fibrous and necrotic tissue to 5 voxels (0.09 mm). The reason for this difference is that the ground truth for calcification was based on the registration of μCT and MRI, which was assumed to be more accurate than that of histology and MRI. The erosion was performed in separate 2D slices in the high-resolution domain for classification (indicated in Figure 2.1) and deleted on average 21% of the voxels in the original segmentations. For every voxel the following features were obtained:

- Intensities in the normalized MR images as obtained with the three different sequences, and after convolution with a Gaussian kernel with σ=0.1 mm, σ=0.2 mm and σ=0.3 mm to obtain smoothed images (12 features)

$$IS(\mathbf{x},\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}^3} e^{-\frac{||\mathbf{x}||^2}{2\sigma^2}} * I \tag{2.1}$$

  with I the normalized image, and $\mathbf{x}$=[$x\ y\ z$].
- The gradient magnitude of all smoothed images (9 features)

$$GM = \sqrt{\left(\frac{\partial(IS(\mathbf{x},\sigma))}{\partial x}\right)^2 + \left(\frac{\partial(IS(\mathbf{x},\sigma)}{\partial y}\right)^2 + \left(\frac{\partial(IS(\mathbf{x},\sigma)}{\partial z}\right)^2} \tag{2.2}$$

- The Laplacian of all smoothed images (9 features)

$$L = \frac{\partial^2(IS(\mathbf{x},\sigma))}{\partial x^2} + \frac{\partial^2(IS(\mathbf{x},\sigma))}{\partial y^2} + \frac{\partial^2(IS(\mathbf{x},\sigma))}{\partial z^2} \tag{2.3}$$

- The shortest (Euclidean) distance to the lumen and to the outer vessel wall (2 features)

This resulted in a total set of 32 features. Calculation of the smoothed, gradient magnitude and Laplacian features was done in the original 3D MRI volumes before registration. The distances were calculated in-plane in the 2D slices after registration. Classification was performed with a linear discriminant classifier which finds the optimal linear boundaries between the classes under the assumption that the data is normally distributed with equal covariance matrices for the three classes. The discriminant function is as follows [76]:

$$\rho_k(\mathbf{x}) = \mathbf{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_k - \frac{1}{2}\boldsymbol{\mu}_k^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_k + \log\pi_k \tag{2.4}$$

Here k are the classes, $\boldsymbol{\Sigma}$ the covariance matrix, $\boldsymbol{\mu}_k$ the class means, π_k the class prior probabilities, and $\mathbf{x}$ the feature vector to classify.

2.2.4 Evaluation

For both the 3D and 2D registration approach, the accuracy of the final registration of histology with *ex vivo* MRI was evaluated. We calculated the distance between the annotated landmark points, and the distance between corresponding vessel wall contours. For the contour distance, the shortest Euclidean distance from each point on the lumen or outer vessel wall contour in the MR images to the corresponding contour in histology was taken, and the other way around. The mean of all these distances was calculated for each slice. The registration of *ex vivo* MRI with μCT was evaluated based on the landmark distance after rigid transformation of the μCT.

The classifiers were evaluated using leave-one-out cross-validation in which one specimen per time was kept out of the training set and used for testing. One percent of the voxels in the remaining 11 specimens, was randomly selected to form the training set. Due to the large number of voxels in the images used for classification, an average of 5000 voxels of the smallest component (calcification) were still present in the training set. Both classifier training and testing was performed using the Matlab toolbox “prtools" [50], version 4.1.9

Classifier performance was evaluated on basis of the voxelwise classification accuracy averaged over the 12 specimens. For feature evaluation, eight combinations of features were evaluated in the complete 3D-registered dataset. Firstly the original MRI intensities as used previously [37, 38, 154] and the original images combined with either smoothed intensities, distances (as in [119]), or gradient magnitude and Laplacian images were considered to study the effect of the new features separately on their added value. In addition, a combination of original and smoothed intensities and distances was evaluated, and this combination with the gradient magnitude features, the Laplacian features, or both added. Differences in accuracy between classifiers using these different feature sets were tested for significance with a Friedman test, a paired nonparametric test for multiple comparisons. Individual comparisons were done using the Tukey-Kramer method.

To study whether accuracy is affected by 2D or 3D registration and slice selection, classification was performed in the three different datasets: the complete 3D-registered set with 11±4 slices per specimen, the 2D-registered dataset with 3 slices per specimen, and the selection of the same 3 slices per specimen from the 3D-registered dataset. Classification was performed using all features and the same leave-one-out procedure as mentioned before. Friedman analysis was again used to test for significant differences.

2.3 Results

2.3.1 Registration accuracy

For 3D registration, the distance between the annotated landmarks in histology and *ex vivo* MRI after all registration steps was 1.3±0.7 mm. The distance between lumen contours was on average 0.5±0.3mm and between outer vessel wall contours 0.3±0.2mm. Using 2D registration, the distance between landmarks was 1.8±1.7 mm, between lumen contours 0.4±0.2 mm and between outer vessel wall contours 0.3±0.3 mm. The distance between the landmarks on *ex vivo* MRI and μCT after applying the rigid transformation was 0.3±0.7 mm.

2.3.2 Performance of features

The average leave-one-out accuracy, sensitivity and specificity for different combinations of features are shown in Table 2.2. To compare, random class assignment according to the class priors (4% for calcification, 63% fibrous and 33% necrotic tissue) would result in an accuracy of 51%. A combination of the three original image intensities yielded an accuracy of 60.8±16.1%. Smoothed intensities, distances and the gradient magnitude together with Laplacian images all improved accuracy when separately added to the original intensities, although none of them statistically significant. When all features were included, the accuracy was the highest (73.3±6.3%), and significantly better than classifiers that have been studied before which used only original intensities, or original intensities and distances (Figure 2.2).

Figures 2.3 and 2.4 show two examples of a slice that was classified using only the original intensities, original intensities and distances, and all features. In Figure 2.3, the classification accuracy increased from 62.3% to 74.9% using all features instead of only original intensities. In Figure 2.4, the accuracy increased from 30.5% to 61.6%. The main improvement in both examples, and in most unshown cases, is due to a more accurate detection of lipid-rich necrotic tissue.

2.3.3 Performance of best classifier

The performance of the best classifier (i.e. including all features) is evaluated in more detail. Table 2.3 shows the confusion matrix of the different plaque components. As also shown in Table 2.2, calcification can accurately be classified with high sensitivity (81%) and specificity (97%). The specificity for fibrous tissue is lower (60%, with a sensitivity of 85%) while for necrotic tissue the sensitivity is lower (52%, with a specificity of 89%). This indicates that the differentiation between fibrous and necrotic tissue is more difficult, and, as Table 2.3 shows, may lead to the misclassification of necrotic tissue as fibrous tissue. The accuracies per class, considering the other two classes as one, are 96% for calcification, 76% for fibrous tissue and 77% for necrotic tissue. The Receiver-Operating curves (Figure 2.5) also show that a very good sensitivity and specificity can be achieved for calcification (area under the

Table 2.2: Classification results for different feature combinations. The overall voxelwise classification accuracy is given as an average over the 12 specimens. Sensitivity and specificity for the 3 components are calculated after combining the 12 confusion matrices and given as one value for all specimens together. The last column indicates which feature sets had significantly different results (Friedman analysis, $p<0.05$). I = original intensities (3 features), IS = smoothed intensities (9 features), D = distances (2 features), GM = gradient magnitude (9 features) and L = Laplacian (9 features).

Feature set	Accuracy (%)	Calcification (Sens, Spec)	Fibrous (Sens, Spec)	Necrotic (Sens, Spec)	Friedman ($p<0.05$)
1 - I	60.8±16.1	81, 96	91, 23	10, 94	5, 6, 7, 8
2 - I, IS	68.1±7.8	79, 96	83, 52	42, 87	
3 - I, GM, L	70.1±6.4	80, 96	82, 56	48, 86	
4 - I, D	70.5±6.8	79, 96	84, 55	46, 88	8
5 - I, IS, D	72.6±6.4	78, 97	85, 59	51, 88	1
6 - I, IS, D, L	72.9±6.6	79, 97	85, 59	52, 88	1
7 - I, IS, D, GM	73.0±6.2	80, 96	85, 60	52, 88	1
8 - All 32 features	73.3±6.3	81, 97	85, 60	52, 89	1, 4

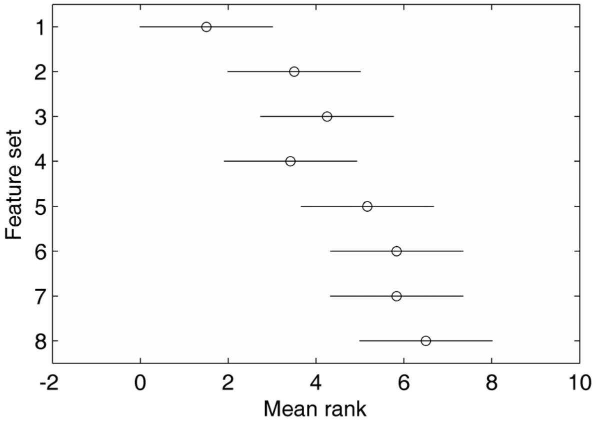


Figure 2.2: Friedman analysis. For each feature set, numbered in Table 2.2, the mean rank is shown.

curve (AUC) of 0.98), with lower areas under the curve for fibrous (AUC of 0.82) and necrotic tissue (AUC of 0.81).

Figure 2.6 shows correlations between the fractions of the different plaque components found in the ground truth and by classification. Values for Spearman correlation coefficients (ρ) were reasonable to good (0.86 for calcification, 0.71 for fibrous tissue and 0.72 for necrotic tissue). On average, the percentage of calcification in the ground truth is 2.3±2.4% lower than in the classification result. Fibrous tissue is 7.7±11.6% overestimated and necrotic tissue 10.1±12.2% underestimated. For illustration, Figure 2.7 shows four examples. Figure A shows a slice with high classification accuracy (83.6%), B and C with average accuracy (74.0% and 69.2%) and D with low accuracy (52.4%). These examples again show the accurate detection of calcification and the more difficult separation between fibrous and necrotic tissue.

Table 2.3: Confusion matrix for classification using all features, summed over all specimens.

		Classification result		
		Calcification	Fibrous	Lipid-rich necrotic
Ground truth	Calcification	3.6%	0.5%	0.4%
	Fibrous	2.0%	53.5%	7.3%
	Lipid-rich necrotic	1.3%	14.4%	17.0%

2.3.4 2D versus 3D registration

The classification results in the 2D-registered and 3D-registered datasets that contained three slices per specimen are shown in Table 2.4. The overall accuracy in these datasets is slightly lower (71.9±11.3% using 3D registration and 70.1±9.0% using 2D registration) than when using the complete 3D-registered dataset. The differences are, however, not statistically significant. The smaller datasets show a slightly higher sensitivity for necrotic tissue and a lower sensitivity for fibrous tissue. The average rotation angle needed to align the MRI volumes with the 3D histology stack was 7±4°.

Figures 2.8 and 2.9 show classification of two slices that are part of all three datasets. While

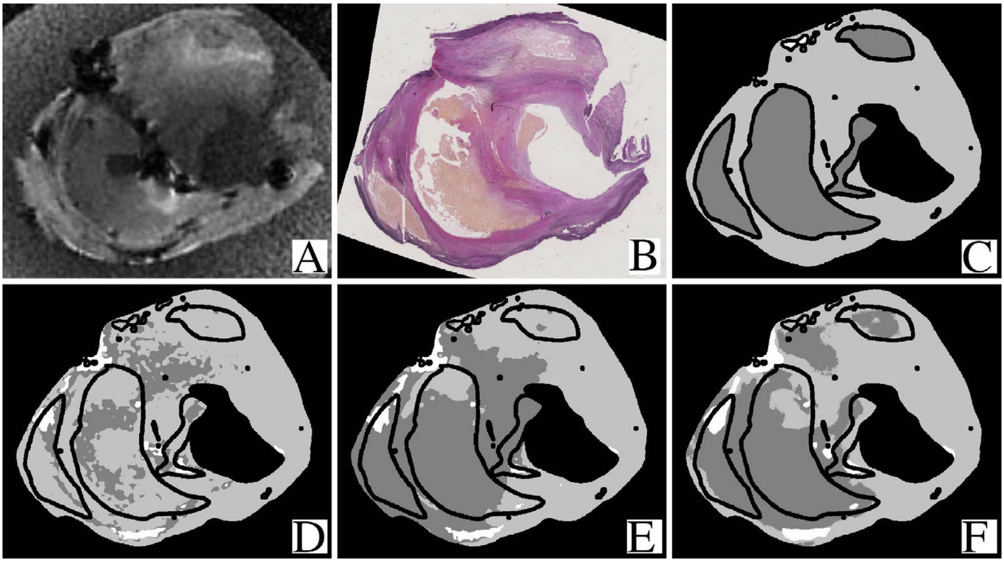


Figure 2.3: Example of classification in one slice. A shows the 3D-T1w MRI slice, B the registered histology and C the ground truth. The classification results are shown using only the three original intensities (D, 62.3% accuracy), original intensities with distances (E, 67.3% accuracy), and all 32 features (F, 74.9% accuracy). In C-F, white represents calcification, light grey fibrous and dark grey lipid-rich necrotic tissue. Black lines and spots are due to the erosion, and not included in the analysis.

the matched 2D MRI slice and the corresponding slice from the 3D-registered MRI volume in Figure 2.8 may look similar, the MRI volume was rotated for 10.7° for the 3D registration. This example shows indeed the highest classification accuracy after training on the complete 3D set, followed by the selection from the 3D set and the 2D set. Due to differences in non-rigid deformation, the ground truth images after 2D and 3D registration may differ slightly. As the μCT volume was rotated together with the MRI, the location of calcification varies. Figure 2.9 shows an example in which the rotation of the MRI volume for 3D registration was larger (13.4°). This example also shows the misclassification of fibrous and necrotic tissue as calcification caused by the presence of air in the specimen which results in black spots in the MRI.

Table 2.4: Classification results for the three different datasets. The overall voxelwise classification accuracy is given as an average over the 12 specimens. Sensitivity and specificity for the 3 components are given as the total sensitivity and specificity across all 12 specimens combined. Differences are not significant.

Dataset	Accuracy (%)	Calcification (Sens, Spec)	Fibrous (Sens, Spec)	Necrotic (Sens, Spec)
Complete 3D set	73.3±6.3	81, 97	85, 60	52, 89
Selection from 3D set	71.9±11.3	81, 98	82, 62	57, 85
2D set	70.1±9.0	70, 96	78, 66	61, 83

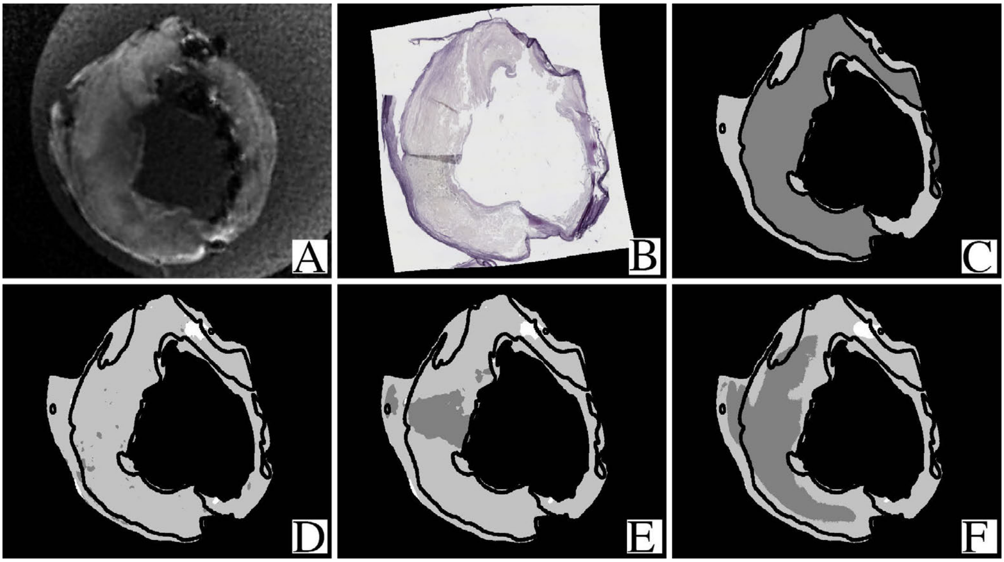


Figure 2.4: Example of classification in one slice. A shows the 3D-T1w MRI slice, B the registered histology and C the ground truth. The classification results are shown using only the three original intensities (D, 30.5% accuracy), original intensities with distances (E, 41.0% accuracy), and all 32 features (F, 61.6% accuracy). In C-F, white represents calcification, light grey fibrous and dark grey lipid-rich necrotic tissue. Black lines and spots are due to the erosion, and not included in the analysis.

2.4 Discussion

We presented an automated method for voxelwise classification of carotid plaque composition in *ex vivo* MRI. Classification accuracy was significantly improved by features that describe local image structure and the location within the vessel wall, in addition to image intensity. Using 3D registration of MRI to histology with as many slices included for training and evaluation as possible (11±4), gave slightly but not significantly better results than when only three slices per specimen were included, or when 2D registration was used to match MRI and histology. The method showed high classification accuracy for calcification (96%) and reasonable accuracy for fibrous (76%) and lipid-rich necrotic tissue (77%).

Four types of features were used in addition to original intensities: the distances to the lumen and outer vessel wall, and smoothed, gradient magnitude and Laplacian images at several scales. The positive effect of including distance information has been shown previously [119]; it is relevant because the necrotic core is usually located in the center of a plaque, surrounded by fibrous tissue. Because smoothed images reduce noise and include information from neighbouring voxels, they may therefore improve performance. Gradient magnitude and Laplacian features are useful to detect small structures; in this study their main effect was an improved sensitivity to calcification detection. Although the improvement of adding gradient magnitude and Laplacian images was small as calcification was already accurately classified, only the classifier that used all features was significantly better than the classifier that used only the original intensities and distances.

The component that is detected with the highest accuracy is calcification, which appears dark in MRI compared to the rest of the plaque. There is a larger overlap in intensity between fibrous and necrotic tissue in the MR images, which makes the differentiation be-

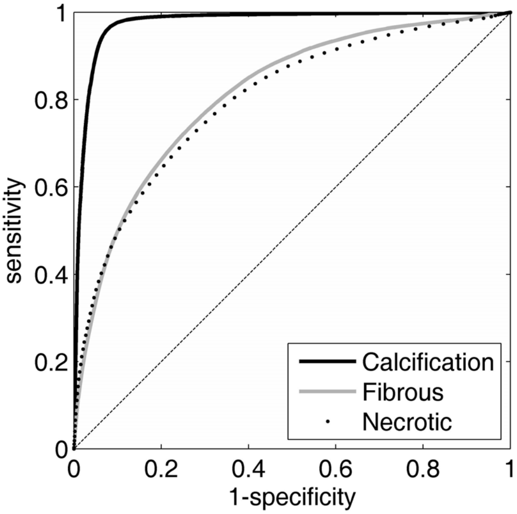


Figure 2.5: ROC curves for the three components studied. The area under the curve for calcification is 0.98, for fibrous tissue 0.82 and for lipid-rich necrotic tissue 0.81.

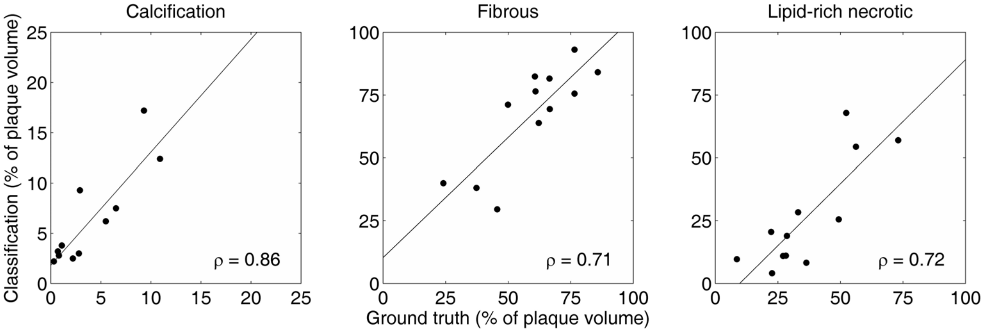


Figure 2.6: Correlation between the relative size of plaque components in the ground truth and classification results. Values for Spearman's ρ are 0.86 for calcification, 0.71 for fibrous tissue and 0.72 for lipid-rich necrotic tissue.

tween these components more difficult. Another reason for the difference in accuracy is that registration of *ex vivo* MRI to μCT is slightly more accurate than that of *ex vivo* MRI to histology due to deformations that occur during histological sectioning. Histology yielded the ground truth for fibrous and lipid-rich necrotic tissue and μCT for calcification. We accounted for small registration errors by erosion of the ground truth areas. The erosion was a little smaller (0.17 mm between fibrous and necrotic regions) than the average contour distance (0.5±0.3 mm for lumen and 0.3±0.2 mm for the outer wall using 3D registration, and 0.4±0.2 mm for lumen and 0.3±0.3 mm for the outer wall using 2D registration). Although misregistration may have affected results in areas with larger errors, most smaller misregistrations were accounted for and visual inspection does not suggest misregistration to be a problem. Differentiation between fibrous and necrotic tissue is important to separate vulnerable from stable plaques. The Spearman's rank correlation coefficients indicate that when the plaques are ranked on the relative volume per component, the result corresponds moderately to well with the ground-truth ranking (Figure 2.6). It is therefore possible to use these plaque component sizes in studies that relate plaque composition to prognosis or outcome and possibly use it for the detection of vulnerable plaques.

3D registration of MRI to histology seemed to result in slightly, although not significantly, higher classification accuracies than when using 2D registration. The rotation angles for 3D registration were quite small (7±4°). Although in this study the difference in classification using either 3D registration or 2D registration was small as well, we believe that 3D registration can provide more accurate training data. This will especially be the case for datasets in which the angle between the slicing direction of the histology specimen and the MRI scan direction is larger, for example in *in vivo* studies. In addition, registration of the complete volume at once is more consistent, and reduces the chance that a single slice will be misregistered. The difference found when many slices or when a selection of slices was used, may be explained by the fact that the selected slices had a slightly different composition (57% fibrous and 38% necrotic tissue in the subset from the 3D-registered set and 55% fibrous and 40% necrotic tissue in the 2D-registered set, against 63% and 33% in the complete dataset). Classification accuracy may therefore be higher in the complete dataset because it contains slices that consist of more fibrous tissue, which may be easier to classify.

These results are comparable to those of previous studies that performed voxelwise plaque classification in *ex vivo* MRI [37, 38, 154]. These studies used only the original MRI intensities and had resolutions slightly lower than in the present study. When 8 contrast weightings were used Clarke et al. reported an overall accuracy of 73.5% [38], or 78% when either these 8 scans or a combination of T1w, T2w and diffusion weighting (Dw) was used [37]. In Clarke's study a combination of T1w, T2w and PDw, as used in the present study, resulted in an accuracy of 67% [37]. Another study that misclassified more than 40% of the voxels when five components were distinguished and that misclassified 20-30% when fibrous and loose connective tissue, and hemorrhage and necrotic tissue were combined, also yielded slightly better results when PDw images were replaced with Dw images [154]. Our study did not include Dw images; classification accuracies might therefore improve when this contrast weighting is added. Using only the original intensities, we obtained an accuracy of 61%, which is slightly lower than in previous studies. Differences with these previous studies may be related to image quality, MRI protocols, the image slices selected, and possibly other factors. However, we expect the added value of the studied features to be valid on other data as well. As shown in Table 2.1, MRI scanning parameters varied between specimens in our protocol, but we did not observe a correlation between deviation from the mean scanning parameters and classification accuracy. The studies by Clarke et al. classified fibrous, loose connective, calcified and necrotic tissue [38] and fibrous, necrotic, calcified and hemorrhagic tissue [37]. We decided not to differentiate between fibrous and loose connective tissue, as they are both stable components and often occur mixed. This makes it difficult to distinguish them in histology, and combining them therefore improved the reliability of our ground truth. Hemorrhage was not included in our study as this plaque component cannot be identified on histological cross-sections with Elastica von Gieson stains. If a ground truth for hemorrhage is present, it can easily be added to the framework.

2.5 Conclusion

We presented a new system for the voxelwise classification of calcification, fibrous and lipid-rich necrotic tissue in *ex vivo* MRI of carotid specimens. Our approach uses 3D registration of MRI with a histology stack to train and evaluate classifiers. Features used for classification are original MRI intensities as well as smoothed intensities, gradient magnitude images, Laplacian images, and distances to the lumen and outer vessel wall. This yields results that are significantly better than when only original intensities and distances are used. The classification results slightly improved by 3D registration compared to 2D registration.

Acknowledgements
Kim van Gaalen is acknowledged for her help with the histological processing and creating the 3D histology.

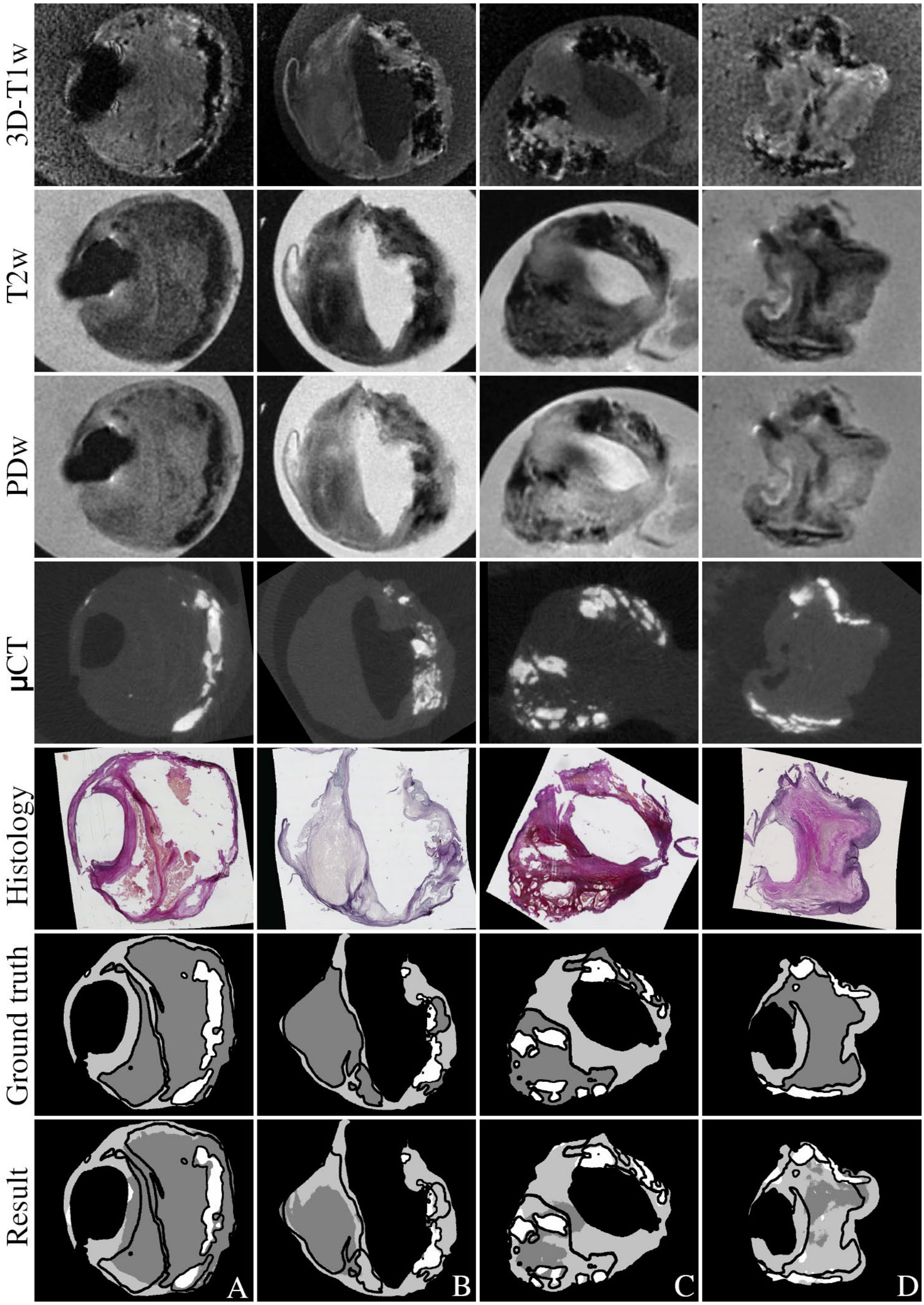


Figure 2.7: Four examples of slices classified using all features. The slice in A has a high classification accuracy (83.6%), the slices in B and C an average accuracy (74.0% and 69.2%), and the slice in D has a low classification accuracy (52.4%). In the ground truth and results, calcification is white, fibrous tissue light grey and lipid-rich necrotic tissue dark grey. For visualization, the MRI and μCT images have been individually scaled relative to the minimum and maximum intensity value to obtain an optimal contrast per image.

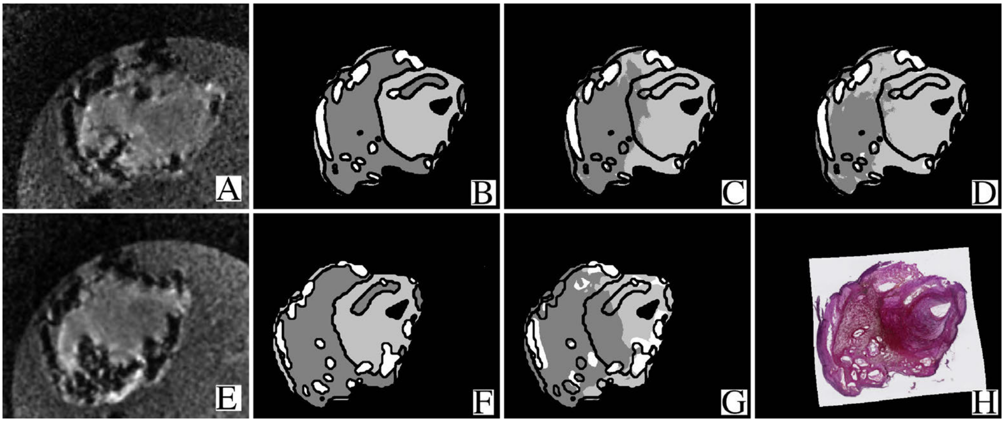


Figure 2.8: Example showing the same histology slice (H) classified in the three different datasets. Image A was matched to this histology slice using 3D registration; image E was matched using 2D registration. B (3D) and F (2D) give the corresponding ground truth. C shows the classification result using the complete 3D-registered set for training (accuracy 76.5%), D using a selection from the 3D-registered set (74.5%), and G using the 2D-registered set (61.6%).

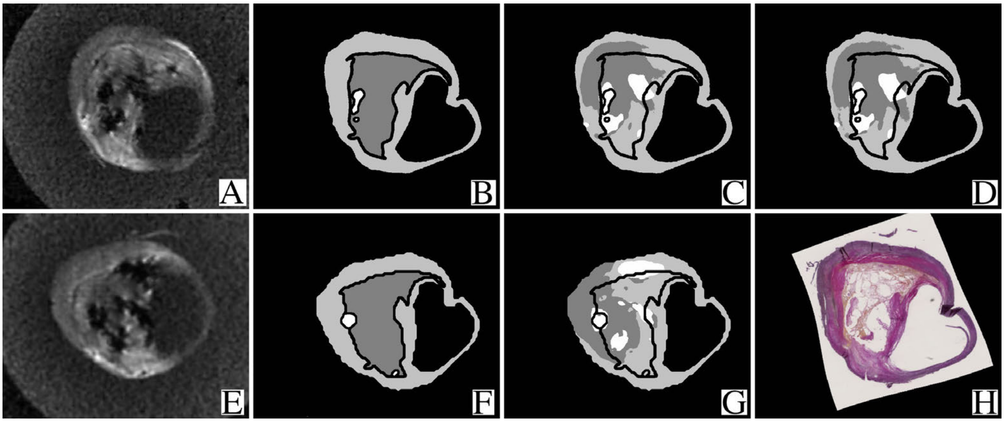


Figure 2.9: Example showing the same histology slice (H) classified in the three different datasets. Image A was matched to this histology slice using 3D registration; image E was matched using 2D registration. B (3D) and F (2D) give the corresponding ground truth. C shows the classification result using the complete 3D-registered set for training (accuracy 66.2%), D using a selection from the 3D-registered set (67.2%), and G using the 2D-registered set (52.2%).

Chapter 3

Atherosclerotic plaque component segmentation in combined carotid MRI and CTA data incorporating class label uncertainty

This chapter has been published as:
A. van Engelen, W.J. Niessen, S. Klein, H.C. Groen, H.J.M. Verhagen, J.J. Wentzel, A. van der Lugt and M. de Bruijne, Atherosclerotic plaque component segmentation in combined carotid MRI and CTA data incorporating class label uncertainty, *PLOS ONE*, 2014

* * *

Abstract
Atherosclerotic plaque composition can indicate plaque vulnerability. We segment atherosclerotic plaque components from the carotid artery on a combination of *in vivo* MRI and CT-angiography (CTA) data using supervised voxelwise classification. In contrast to previous studies the ground truth for training is directly obtained from 3D registration with histology for fibrous and lipid-rich necrotic tissue, and with μCT for calcification. This registration does, however, not provide accurate voxelwise correspondence. We therefore evaluate three approaches that incorporate uncertainty in the ground truth used for training: I) soft labels are created by Gaussian blurring of the original binary histology segmentations to reduce weights at the boundaries between components, and are weighted by the estimated registration accuracy of the histology and *in vivo* imaging data (measured by overlap), II) samples are weighted by the local contour distance of the lumen and outer wall between histology and *in vivo* data, and III) 10% of each class is rejected by Gaussian outlier rejection. Classification was evaluated on the relative volumes (% of tissue type in the vessel wall) for calcified, fibrous and lipid-rich necrotic tissue, using linear discriminant (LDC) and support vector machine (SVM) classification. In addition, the combination of MRI and CTA data was compared to using only one imaging modality. Best results were obtained by LDC and outlier rejection: the volume error per vessel was 0.9±1.0% for calcification, 12.7±7.6% for fibrous and 12.1±8.1% for necrotic tissue, with Spearman rank correlation coefficients of 0.91 (calcification), 0.80 (fibrous) and 0.81 (necrotic). While segmentation using only MRI features yielded low accuracy for calcification, and segmentation using only CTA features yielded low accuracy for necrotic tissue, the combination of features from MRI and CTA gave good results for all studied components.

3.1 Introduction

Atherosclerotic disease of the carotid artery is common in the elderly population, and is a major cause of cerebral ischemia [65, 83]. The underlying mechanism is the rupture of atherosclerotic plaque with subsequent embolisation of thrombus and/or plaque material in the cerebral circulation. Clinical manifestations and fatal outcomes are most often associated with plaques of American Heart Association (AHA) type IV, V and VI [184]. On MRI these are characterized by presence of a lipid or necrotic core (LRNC) and possibly calcifications (type IV-V) or a possible surface defect, hemorrhage or thrombus (type VI) [29]. These characteristics as found in carotid histology have been related to recent symptoms [167, 168], and measurements of tissue components from MRI have been related with future events [160, 188]. Prevention of (recurrent) cerebral ischemia is the goal of pharmacological or surgical treatment. Currently the decision for surgical treatment such as carotid endarterectomy or carotid artery stenting is based on the degree of stenosis, but incorporating non-invasive measures of plaque composition is expected to improve the selection of patients that will benefit from surgical intervention [26, 143, 242].

Non-invasive identification of different plaque components is possible both with magnetic resonance imaging (MRI) [77, 158, 243] and CT-angiography (CTA) [45, 237]. Manual component segmentation and quantification in MRI is time-consuming and subject to inter- and intraobserver variability [158, 187]. Automated segmentation methods that are accurate and robust are therefore essential to perform large scale studies that can determine the clinical relevance of plaque composition, and to be able to incorporate these measures into daily clinical practice if this is deemed advantageous. In this paper we perform automatic segmentation of plaque components using a combination of MRI and CTA images and evaluate the advantage of combining those imaging modalities.

Automated methods that segment plaque components have been developed previously, but these show some limitations in the accuracy for different components, and use possibly inaccurate or biased training and evaluation methods. Considering the accuracy for different plaque components, in MRI good results have generally been obtained for quantification of fibrous tissue and LRNC, but except for the results reported in [119] a low accuracy for calcification has been found [82, 219]. CTA on the other hand, provides a good estimation of calcium volume, while the differentiation between LRNC and fibrous tissue is more challenging due to a large overlap in Hounsfield values [45, 237]. A combined analysis of MRI and CTA may be beneficial for accurate quantification of all plaque components [74, 132].

The use of possibly inaccurate or biased training and evaluation methods results from the difficulty to obtain an accurate ground truth. Supervised pattern classification is commonly used as part of segmentation methods, and voxelwise classification has also been applied to successfully segment plaque components from MRI [82, 119, 219]. These techniques require a known voxelwise ground truth for classifier training. The ground truth may be obtained from manual segmentations [219], but these may be inaccurate due to overlapping intensities between classes and inter-observer and intra-observer variability [82, 119]. Histology sections are considered to be more objective [124], but it is difficult to accurately align these with *in vivo* scans due to tissue deformations that occur during surgical plaque excision and histology processing. Histology-guided manual annotations have been used as well [82, 119], but may introduce a bias toward the *in vivo* scan data [136]. In this paper we choose to use the more objective information directly obtained from histology, while trying to account for misregistration during classifier development.

Image registration between histology and *in vivo* data is a topic of interest in many applications [24, 41, 230], and using the registration with histology as a ground truth for *in vivo* pattern classification is a challenging problem [7, 225]. For atherosclerotic plaque, image registration between histology and *in vivo* data has mostly been done by manually selecting corresponding slices followed by rigid 2D registration [45, 82, 119, 237]. Non-rigid 3D registration has also been used, to allow rotation of the *in vivo* image orientation with respect to the histology slicing direction, and to compensate for in-plane deformations in histology [68]. Although this does allow for the correct rotation angle, it remains difficult to obtain voxelwise correspondence for the vessel wall and plaque components due to the large deformations that occur owing to plaque excision and histology processing.

Another approach to handle registration inaccuracies when registered data is used to train a classifier, is to account for inaccurate sample labels during the training phase. Several ways to cope with inaccurate labels have been proposed. One approach is to detect outlier samples and reject those samples from the training set. An overview of methods for outlier rejection is given by Hodge and Austin [81]. Another approach is to adjust the weight or label of samples with an uncertain label. Bouveyron and Girard [22] used prior clustering to detect samples with inconsistent labels and took these inconsistencies into account during supervised modeling. Prior clustering has also been used to create fuzzy labels that indicate a membership probability for each class [118]. In this way outlier samples get a low membership value for the class they belong to according to the hard label. Thiel [200] showed that classifiers based on such 'soft' labels are robust against label noise by artificially adding different levels of noise to soft labels. The approaches above use distances in feature space to determine soft labels. In our case we have additional knowledge on the probability that labels are accurate. At the border between plaque components errors are more likely to occur than in the center because of misregistration between histology and *in vivo* images. In addition, locations where the histology and *in vivo* images align well provide more accurate labels than locations that are less well registered. We evaluate two approaches that use this information to modify the sample labels and/or weights and compare to using the original hard labels and a standard way of (Gaussian) outlier rejection. In addition, registration between different MRI sequences, and between CTA and MRI, is important for accurate classification. We will present an approach for these registrations.

In this paper we perform plaque component segmentation in *in vivo* imaging data. We combine MRI and CTA scans to differentiate between calcification, fibrous tissue and lipid-rich necrotic tissue. The main contributions of this paper are 1) the evaluation of different approaches for training on histology data which account for registration errors, 2) the combination of MRI and CTA imaging features for plaque characterization and the evaluation of their performance, and 3) optimization of a 3D registration framework to match *in vivo* MRI and CTA with histology. Together these steps present a framework for quantification of plaque components in *in vivo* data, by training on registered histology. A preliminary version of this paper has been presented previously at a conference [214]. The current paper presents more ways of handling registration accuracy during training, has a more elaborate evaluation and discussion, and includes a comparison between MRI and CTA.

3.2 Materials and Methods

This section is structured as follows. After the Ethics statement we describe the data in Section 3.2.2, which consist of 13 arteries (13 patients) that are all imaged with corresponding histology, *ex vivo* MRI, μCT and *in vivo* MRI and CTA. Image registration of the *in vivo* images and histology sections, in order to obtain a ground truth to train the segmentation method, is described in Section 3.2.3. Section 3.2.4 subsequently describes the design of the classifiers, including the different ways of handling registration inaccuracies for training. Finally, in Section 3.2.5 the experiments for evaluation are presented. Final registered data of all subjects (CTA, MRI, histology, labeled ground truth and *in vivo* wall segmentation) will be made available upon request.

3.2.1 Ethics statement

This study was approved by the Medical Ethical Committee of the Erasmus Medical Center. Written informed consent was obtained from all subjects.

3.2.2 Data

Fifteen patients (all male, age 68±9 years) who were scheduled for carotid endarterectomy (CEA) were selected for this study and gave informed consent. Nine had an ischemic stroke, five had a transient ischemic accident and one was asymptomatic. A subset of this data has previously been used for the development of a 3D registration framework [68] and to develop a segmentation method on *ex vivo* MRI [Chapter 2]. Due to incomplete imaging data and a low quality of histology two patients were excluded, leaving thirteen datasets for the analysis. Before CEA, patients underwent *in vivo* MRI (Signa Excite (3 Tesla), GE Healthcare, Milwaukee, USA) and CTA (Sensation 16 (n=4)/Sensation 64 (n=9), Siemens, Erlangen, Germany) scanning. MRI was made one day prior to CEA, and CTA 38±26 days earlier. We used four MRI scans that were made before contrast administration (2D-T1w, 2D-PDw, 2D-TOF and 3D-T1w), and one 3D-T1w scan 4.6±3.4 minutes after intravenous administration of gadofosveset (Vasovist, 0.03 mmol/kg body weight, Bayer Schering Pharma AG). Details are provided in Table 3.1. Due to the better performance of contrast-enhanced T1w scans to differentiate plaque tissues compared to T2w scans, no T2w scans were used [28, 231, 246]. CTA images were made with a standardized contrast-enhanced protocol [43] and had a resolution of 0.27±0.05 mm in-plane with a slice thickness of 0.9±0.1 mm and a slice distance of 0.5±0.1 mm. After registration of the MRI and CTA scans (Section 3.2.3) manual annotations of the vessel wall were made on the registered *in vivo* scans. Annotations were based on a combination of CTA, PDw MRI and postcontrast T1w MRI, with visual inspection of the other in vivo MR sequences.

As previously published [68], to facilitate registration of the *in vivo* data to histology, *ex vivo* MRI (3D-T1w Gradient Echo, 0.1×0.1×0.1 mm, Signa Excite, GE Healthcare) and μCT scans (18×18×18 μm, Skyscan 1072, Skyscan, Belgium) of the excised plaque were made. In addition photographs of the specimen were taken every 1-mm interval during histology slicing, called 'block-face' images (15±1×15±1 μm). Histology sections were taken every 1-mm interval (1.8×1.8 μm) and stained with Elastica von Gieson staining (Merck, Germany). To obtain ground truth segmentations the vessel wall was manually segmented in histology, and divided into fibrous and lipid-rich necrotic regions. The ground truth for calcification was obtained by thresholding the μCT at a fixed value for all scans [Chapter 2]. Based on histology quality, 11±4 histology slices with registered *in vivo* images were included per subject (range 3-17 slices).

Table 3.1: MRI settings. FSE = Fast spin echo, TOF = Time of flight, GRE = Gradient echo. *Same settings were used pre- and postcontrast.

	Repetition time (ms)	Echo time (ms)	Flip angle	In-plane resolution (mm)	Slice thickness	Slice distance
2D-T1w FSE	425±77	12.1±1.1	90°	0.41±0.07	1.5	1.5
2D-PDw FSE	4635±284	17.2±1.9	90°	0.41±0.06	1.5	1.5
Fast TOF	15.3±1.2	3.4±0.3	40-60°	0.91±0.11	2-3	1.5-2
3D-T1w GRE*	15.3±0.3	3.15	16°	0.61±0.05	0.8-1	0.4-0.5

3.2.3 Image registration

Our registration framework is an extension of the 3D registration of CTA with histology as described by Groen et al. [68]. That method registered CTA to histology image data using the following steps. First, a 3D histology stack is created by non-rigid registration of histology slices to a stack of 'block-face' photographs taken during sectioning. CTA is registered to μCT using isotropic scaling based on annotated landmark points in both imaging modalities, which are mainly calcium spots. To align μCT with *ex vivo* MRI, and *ex vivo* MRI with the 3D histology stack, a rigid transformation based on manually annotated landmarks is applied. Subsequently the 3D histology stack is deformed in-plane to match the *ex vivo* vessel wall, annotated in the *ex vivo* MRI, using a B-spline model [157] that maximizes mutual information (MI) [199, 222] of both image intensity and vessel wall annotations.

We made a number of modifications to this framework. Firstly, we added *in vivo* MRI. All MRI scans were rigidly registered to the postcontrast 3D-T1w scan, and this scan was rigidly registered to the CTA. These registration steps were based on mutual information of image

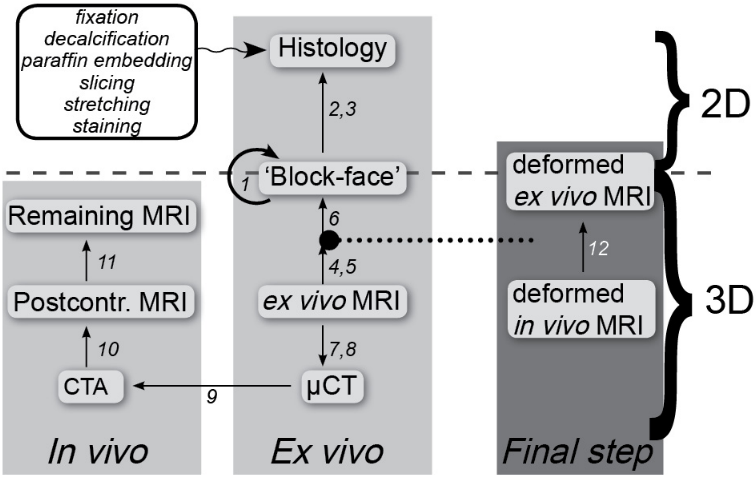


Figure 3.1: Overview of the framework for histology processing and registration. The light gray blocks show registration of *in vivo* MRI and CTA with histology, via *ex vivo* MRI and μCT. The large dot with a dotted line indicates the space to which all images are transformed. After these transformations the registration step in dark gray block is done to directly optimize the registration of *ex vivo* with *in vivo* data. The arrows point from the fixed to the moving image. Numbers in this figure refer to registration steps with a detailed description in Table 3.2

intensity and made use of a mask around the vessel that was annotated in the fixed images (CTA and postcontrast MRI). Secondly, compared with Groen et al. the registration of histology and *in vivo* data was refined in two ways. 1) In the registration of CTA to μCT isotropic scaling was replaced by a thin-plate spline deformation [42], to account for deformations that occur during plaque excision. This was based on landmarks only (7.8±1.6 landmarks per plaque), by fitting an approximating thin-plate spline with relaxation factor of 0.1 [153]. 2) After registration using all previously mentioned steps (light gray area in Figure 3.1), the *ex vivo* MRI was deformed to match the *in vivo* postcontrast MRI. This was done based on maximization of the sum of MI of image intensity, MI of manual segmentations of the lumen and MI of the outer vessel wall with a B-spline model. Similar to the registration of *ex vivo* MRI with histology, a multiresolution scheme with 4 resolution levels and a final B-spline control point spacing of 2 mm was used. Optimization was done using adaptive stochastic gradient descent optimization [101]. The resulting transformation was applied to the ground truth segmentations (histology and μCT), to obtain a better overlap of the vessel wall in the ground truth and the *in vivo* data. The toolbox `elastix` [102] was used for all registrations, in combination with MeVisLab for rigid point-based registration and Python for scripting. A summary of all steps of our modified registration framework is provided in Figure 3.1) and Table 3.2, and a more detailed description and evaluation can be found in [68] and [Chapter 2]. The effect of the two refinements mentioned above is evaluated in Section 3.3.1.

3.2.4 Classifier design

Regions of interest of all *in vivo* MR images were corrected for intensity inhomogeneities using N3 [176], and normalized by setting the mean intensity to 0 and the standard deviation to 100 within these selected regions. A set of 24 image features was calculated for each voxel: the intensities in the normalized 3D-T1w (pre- and postcontrast), T1w, PDw and TOF images, these images blurred with a Gaussian filter (σ=1 mm), the gradient magnitude and Laplacian at the same scale, the original CTA intensity, the Euclidean distances to the lumen and outer vessel wall, and the product of these distances. Intensity, first and second order derivatives and distances have previously proven to be effective [119, 219, Chapter 2]. The product of the two distances was added to enable a linear separation between LRNC and fibrous tissue, which better prevents the lipid-rich necrotic core from touching the lumen or outer vessel wall border. Together these distance features represent both wall thickness and the voxel location relative to the lumen and outer wall. For training the distances were based on the deformed histology segmentation, for testing on the distance to the manual *in vivo* contours. All images (ground truth and features) were resampled to 0.25×0.25 mm in-plane using cubic B-spline interpolation, such that they had a resolution in the order of the *in vivo* CTA.

To account for registration inaccuracies, which lead to inaccurate training labels, we compared three approaches:

1. Uncertainties in the ground truth were taken into account by two mechanisms. First, the binary ground truth segmentations (calcification (C), lipid-rich necrotic tissue (LRNC) and fibrous tissue (F)) were blurred with a Gaussian filter with standard deviation σ (G_σ), followed by normalizing the sum of the three components to 1. This creates soft labels that indicate a probability of belonging to each of the three components, where points close to component boundaries get a similar, lower, probability for multiple components. Second, since the reliability of the labels depends on the

Table 3.2: Settings for the different registration steps.*

	Fixed image	**Moving image**	**Def.model**	**Information**	**Comp. time**	**Manual time**
1	Stacking of block-face images, by registering each slice to its adjacent slice		Rigid	Landmarks	Few sec per slice	~2-3 min per slice
2	Block-face	Histology	Rigid	Landmarks	Few sec per slice	~2-3 min per slice
3	Block-face	Histology	B-spline	MI of intensity, lumen mask and outer wall mask	~2-3 min per slice	~15 min per slice (including composition in histology)
4*	*Ex vivo* MRI	Block-face	Rigid	Landmarks	~10 sec	~5 min per 3D volume
5*	*Ex vivo* MRI	3D histology	Rigid	MI of intensity, lumen mask and outer wall mask	~1 min	~15 min per 3D volume (histology annotation from step 3)
6	*Ex vivo* MRI	3D histology	in-plane B-spline	MI of intensity, lumen mask and outer wall mask	~3-4 min	- (Uses annotations from steps 3 and 5)
7	*Ex vivo* MRI	μCT	Rigid	Landmarks	0.5-1 min	~5 min per 3D volume
8	*Ex vivo* MRI	μCT	Rigid	MI of intensity	0.5-1 min	-
9	μCT	CTA	Thin-plate spline	Landmarks	~10 sec	~5 min per 3D volume
10	CTA	Postcontrast T1w MRI	Rigid	MI of intensity within mask	~0.5 min	~2-3 min per 3D volume
11	Postcontrast T1w MRI	Other MRI images	Rigid	MI of intensity within mask	~0.5 min	~2-3 min per 3D volume
12	Deformed *in vivo* postcontrast T1w MRI	Deformed *ex vivo* MRI	B-spline	MI of intensity, lumen mask and outer wall mask	~5-6 min	~10 min per 3D volume

$*$ For registration the inverse transformation of steps 4 and 5 was applied to the *ex vivo* MRI.
Def. model = deformation model, Comp. time = computation time, MI = Mutual Information

registration accuracy, we estimated registration accuracy. Hereto the Dice overlap between the vessel wall segmentation in histology and registered MRI/CTA was calculated for each slice [49]. To evaluate different degrees of weighting between slices, the soft labels were summed to their corresponding Dice^n to obtain the final soft labels (Figure 3.2A). Here n is an exponent, where with larger n the difference in contribution of slices with low or high Dice overlap becomes larger. The final labels assign a weight to the samples, such that samples close to region boundaries or from slices with a low registration accuracy contribute less to the classifier than samples with a more certain ground truth. In our experiments, we determined the optimum value for σ and n using cross-validation on the training set (Section 3.2.5). In the equation below $w_i(x)$ is the sample weight for class i at voxel x, with M_i the binary mask for class i.

$$w_i(x) = \frac{G_\sigma * M_i(x)}{\sum_i G_\sigma * M_i(x)} \cdot Dice^n \tag{3.1}$$

2. Approach 1 uses the same degree of blurring at all locations to obtain soft labels, and assigns a higher or lower weight to an entire slice based on registration accuracy. In case one part of the section is aligned more accurately than the other, a more local weighting of registration accuracy would be appropriate. To achieve this we calculated for each voxel the Euclidean distance to the lumen and the outer vessel wall, in both histology and the *in vivo* scans. The sample weight $w(x)$ (eq. 3.2) was then defined as a function of weight based on outer wall ($w_{wall}(x)$) and lumen registration accuracy ($w_{lumen}(x)$), where the weight was determined to be 1 for a difference between the histology and *in-vivo* contour of 0 and 0 for a difference of 5 mm, linearly scaled between these values. The ratio of the two weights was determined by the relative distance of the voxel to the lumen and outer wall ($R(x)$):

$$w(x) = \frac{R(x) * w_{wall}(x) + w_{lumen}(x)}{R(x) + 1} \tag{3.2}$$

$$w_{wall}(x) = -0.2 \cdot (|W_{histology}(x) - W_{in-vivo}(x)|) + 1$$

$$w_{lumen}(x) = -0.2 \cdot (|L_{histology}(x) - L_{in-vivo}(x)|) + 1$$

$$R(x) = \frac{L_{histology}(x) + L_{in-vivo}(x)}{W_{histology}(x) + W_{in-vivo}(x)}$$

   where L indicates the distance to the lumen and W the distance to the outer wall. An example is shown in Figure 3.2B.

3. The third approach is Gaussian outlier detection, which excludes samples that are outliers in feature space without taking into account the position of these samples in the original image or the registration accuracy. For each of the three components, 10% of the samples was rejected. A Gaussian target distribution was modelled to the data. The mean and standard deviation for each class were robustly estimated by iteratively reweighing the samples by their distance to the (previously estimated) mean [192]. Outliers can be in misregistered areas for which the image characteristics do not correspond to the class label, but also variations in image intensity for certain scans, or imaging artifacts can be rejected. We choose to reject 10% of the samples in outlier rejection expecting this would be a good balance between not discarding too many samples and at the same time being sure that all outliers are rejected. For an example see Figure 3.2C.

For classification, a linear discriminant classifier (LDC) and support vector machine (SVM) classifier were used. The LDC has been used successfully in previous studies [219, Chapter 2]. The definition is as follows [76]

$$\rho_k(\mathbf{x}) = \mathbf{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k - \frac{1}{2} \boldsymbol{\mu}_k^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k + \log \pi_k, \tag{3.3}$$

where ρ is the posterior probability, k are the classes, $\boldsymbol{\Sigma}$ the pooled covariance matrix, $\boldsymbol{\mu}_k$ the class means, π_k the class prior probabilities, and $\mathbf{x}$ the feature vector to classify. Each sample was assigned to the class with the highest posterior probability. As LDC is a relatively

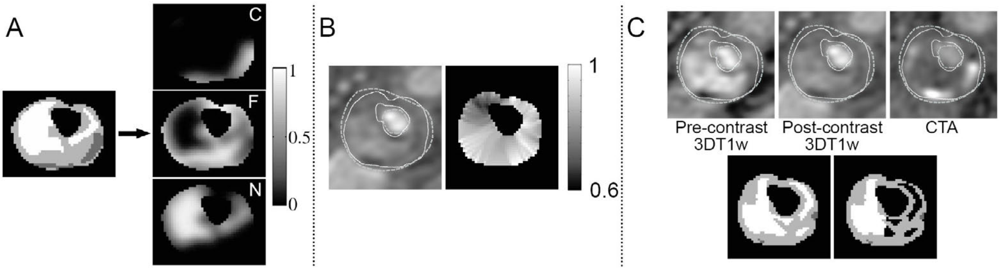


Figure 3.2: Different ways of handling registration accuracy in training. A. Soft labels for each class are derived by blurring the original segmentations. In this example σ=0.5mm and the soft labels of the three classes sum to the Dice overlap between histology and *in vivo* data in each voxel (0.93 in this slice). In the hard segmentation dark gray is calcification (C), light gray fibrous tissue (F) and white lipid-rich necrotic tissue (N). B. Sample weights are determined by the distance between lumen and outer wall contours in histology (white line) and the *in vivo* data (blue dashed line). C. Outlier rejection: Based on 10% outlier rejection on the combination of all 13 vessels, the black areas in the right bottom figure would be rejected as outliers. In this slice mainly lipid/necrotic voxels from the right half of the section are considered outliers in feature space.

simple, non-flexible classifier, we used a support vector machine (SVM) with a radial-basis function (RBF) kernel for comparison. Compared to LDC, SVM is more flexible and it has proven successful in many applications. The classification problem is solved as [31]:

$$\min_{v,v_0,\xi} \frac{1}{2}\mathbf{v}^T\mathbf{v}+C\sum_{i=1}^{l}\xi_i, \quad \text{subject to } y_i(\mathbf{v}^T\phi(\mathbf{x}_i)+v_0)\geq 1-\xi_i,\ \xi_i\geq 0 \tag{3.4}$$

$$\text{with kernel } K(\mathbf{x}_i,\mathbf{x}_j)=\phi(x_i)^T\phi(x_j)=\exp(-\gamma||\mathbf{x}_i-\mathbf{x}_j||^2), \gamma>0$$

with **v** the SVM classifier and **x** the feature vector. C is the penalty parameter of the error term that trades-off between minimizing misclassification and maximizing the margin, γ the kernel radius, ξ the misclassification weight and $y_i \in$ {-1,1} the sample label. The decision boundary is then defined as:

$$f(\mathbf{x})=\mathbf{v}^T\phi(\mathbf{x})+v_0=0 \tag{3.5}$$

New samples are labelled by thresholding $f(\mathbf{x})$ at 0. Multiclass classification was done by combining different one vs one classifiers. For classifier development and evaluation the Matlab toolbox *prtools* [50], and *libsvm* [31] were used.

3.2.5 Evaluation

Leave-one-out experiments were performed in which repeatedly 12 subjects were used for training and the 13^{th} for testing. Classifiers were trained on 10% of all voxels that were within the vessel wall in both the histology segmentation and the *in vivo* wall segmentation, and tested on all voxels within the *in vivo* vessel wall. The same 10% of samples were extracted for training for each experiment.

For labeling approach 1 with ground truth blurring and weighting with Dice^n, σ values equal to 0, 0.25, 0.5, 0.75, 1, 1.5 and 2 mm were evaluated, and n was varied from 0 to 39 with intervals of 3, with a separate cross-validation within the training set of 12 subjects, again by leave-one-subject-out experiments. For each combination of σ and n the absolute error of the calculated volumes (% of the vessel wall) of all three components with respect to the ground truth volumes was averaged over all slices of the 12 subjects, and the σ and n that corresponded to the lowest error were chosen. Using the same approach the penalty parameter c and kernel radius γ were optimized in the training set for all SVM classifications. Here, the features were normalized to have zero mean and a standard deviation of 1.

Classification with LDC and SVM was performed using the four different types of sample labels: 1) original hard labels, 2) labels obtained by ground truth blurring and weighting by the Dice overlap, 3) local weights obtained using the contour distances and 4) Gaussian outlier rejection. All voxels within the *in vivo* segmented vessel wall were classified. The results were evaluated by comparing plaque component volumes as a percentage of the vessel wall with histology, both per subject and per slice. For each classifier, the results per slice were tested for statistical significant differences between the four different approaches. The absolute errors of the three components were averaged as they are strongly related, and compared using Friedman analysis, with post-hoc Tukey-Kramer testing to account for multiple comparisons.

To evaluate performance when classification is based on a single imaging modality, voxel classification was repeated using only MRI and only CTA features, both with and without the distance features. For completeness, also classification using only distance features was evaluated using the same approach. Finally, an experiment was carried out to indicate which features are most relevant for classification. Forward feature selection with LDC accuracy as the evaluation criterion was performed for all three combinations of two components. All voxels of all 13 subjects were used, with their corresponding hard label.

3.3 Results

3.3.1 Registration results

Registering CTA to μCT with a thin-plate spline deformation instead of the isotropic scaling as was used in [68], showed that in cases with large deformations an improved match was obtained (visual inspection). When deforming the *ex vivo* MRI vessel wall to match the *in vivo* MRI vessel wall (dark gray right column in Figure 3.1), the Dice overlap increased from 0.61±0.18 (range 0.14-0.88) to 0.77±0.12 (range 0.31-0.95). Applying this deformation to the histology segmentations increased the Dice overlap between the histology vessel wall and *in vivo* vessel wall from 0.57±0.18 (range 0.11-0.87) to 0.67±0.16 (range 0.22-0.94). The final mean wall distance between histology and *in vivo* data was 0.87±0.63 mm for the lumen and 0.67±0.39 mm for the outer wall. As this error is in the order of several voxels we can assume that voxelwise correspondence was not obtained. Two examples are shown in Figure 3.3.

3.3.2 Segmentation results

The average bias, absolute error and Spearman rank correlations of relative component volumes with respect to the relative volumes in the ground truth (histology and μCT), per subject, are given in Table 3.3. The results show that for LDC good calcification classification was obtained, with errors smaller than 2% and correlation values higher than 0.75. The amount of fibrous tissue was overestimated and the amount of LRNC underestimated, but correlations >0.75 could be obtained too. For LDC, both blurring and Dice weighting, and

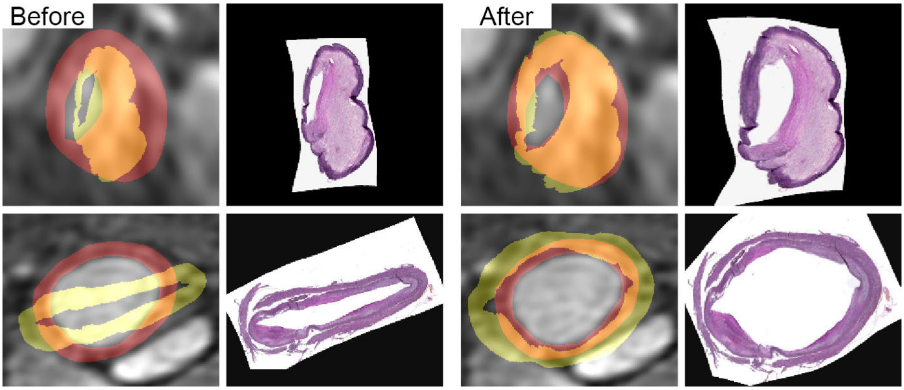


Figure 3.3: Two registered image slices. Examples are presented before and after applying the final deformation step as depicted in the right column of Figure 3.1. In yellow the deformed histology vessel wall is shown, in red the *in vivo* vessel wall overlaid on the postcontrast MRI scan. In the orange regions they overlap. The Dice overlap for the top image increases from 0.59 to 0.86, for the bottom image from 0.28 to 0.44.

outlier rejection decreased the bias, but only outlier rejection reduced the absolute error, for all three components. This error was significantly smaller than for the two methods that use sample weighting, and may be slightly better than using hard labels (p=0.06). Additionally, the improvement in absolute error using outlier rejection was significantly related with the amount of lipid per slice (ρ=-0.23, p<0.01). For slices with higher lipid amounts, the advantage of using outlier rejection was larger than for slices with no or little lipid.

While classification with LDC yielded better correlations than with SVM, SVM yielded a lower bias, and, when weighting by contour distance, a lower absolute error. For SVM the three methods to handle registration errors seem to have a larger effect than for LDC. The correlations for fibrous tissue and LRNC also improved, but the results for calcification deteriorated. The differences in error between the methods with SVM were not significant.

Overall, classification with LDC and Gaussian outlier rejection lead both to a relatively low absolute error and a good correlation with histology for all three components. Therefore, additional visualizations of these results are provided. Scatter plots in Figure 3.4 show the correlation between the relative volume of each tissue component in histology and in the segmentation result. The segmentation results for all slices of one subject are shown in Figure 3.5. The segmentations visually show acceptable spatial fidelity.

The voxelwise accuracy in the overlapping areas of the vessel wall in histology and in *in vivo* data, was 68±6% for LDC and hard labels and 69±6% for LDC after outlier rejection. However, this evaluation is hampered by registration errors and the results should therefore be interpreted with caution.

For the approach with blurring and weighting with LDC the optimal values found for σ (ground truth blurring) were 0.15±0.16 mm (range 0-0.5 mm) with n=23.8±11.5 (range 12-39, ground truth weighting). Thus, no or little blurring of the ground truth segmentations was performed, but high weighting of slices based on registration accuracy was applied. This led to a skewed sample weight distribution with an interquartile range of 0.0001–0.0038–0.0411 (total range 0-0.44). For SVM σ was 0.38±0.35 mm (range 0-1 mm), with n=9.0±6.1

Table 3.3: Segmentation results per subject for different approaches using MRI, CTA and distance features. The results are compared to relative component volumes in histology.∗

Method	Bias (% in result - % in GT)			Absolute error			Spearman (ρ)		
	C	F	LRNC	C	F	LRNC	C	F	LRNC
LDC - H	-0.1±1.5	11.9±10.9	-11.8±10.8	1.0±1.1	14.1±7.5	14.0±7.4	0.91	0.87	0.84
LDC - M1	-0.6±1.9	8.1±14.6	-7.4±13.9	1.5±1.2	14.6±7.4	13.7±7.1	0.77	0.80	0.78
LDC - M2	-0.6±1.7	12.4±11.9	-11.8±11.6	1.1±1.3	15.2±7.4	14.6±7.4	0.85	0.82	0.85
LDC - M3	-0.2±1.4	7.6±13.0	-7.5±12.8	0.9±1.0	12.7±7.6	12.1±8.1	0.91	0.80	0.81
SVM - H	0.5±2.9	-0.5±19.7	0.1±19.3	2.0±2.1	15.7±11.0	15.6±10.4	0.88	0.53	0.58
SVM - M1	1.2±3.9	-4.3±18.1	3.1±17.2	2.6±3.1	12.3±13.6	11.8±12.5	0.68	0.63	0.74
SVM - M2	-1.8±2.0	3.2±11.1	-1.4±10.6	2.1±1.7	9.1±6.7	8.2±6.5	0.63	0.73	0.79
SVM - M3	-1.4±3.8	-1.2±16.9	2.6±14.8	3.0±2.5	12.6±10.6	11.2±9.5	0.43	0.73	0.71

∗ C = calcification, F = fibrous tissue, LRNC = lipid-rich necrotic core, H = hard labels, M1 = blurring and Dice weighting, M2 = weighting by contour distance, M3 = Gaussian outlier rejection.

(range 0-15), and a sample weight interquartile range of 0.014–0.128–0.493 (range 0-1). Using local contour distance the obtained sample weights were 0.91±0.06 (range 0.44-1).

3.3.3 MRI vs. CTA

The experiments to compare performance on MRI and CTA were also performed using LDC and Gaussian outlier rejection. The results using only MRI or CTA are given in Table 3.4. When only MRI features were used, calcification was underestimated and in most cases not detected. Using only the original CTA image, a good correlation for calcification was found, although the volumes were overestimated. Differentiation between fibrous tissue and LRNC was not possible. Adding distance features, however, showed a great improvement. Using the distance features only yielded plausible volume estimates for fibrous tissue and LRNC, however, adding MRI features improved the results even more.

Examples of slices segmented using either MRI or CTA and distance features are shown in Figure 3.6. These show indeed that calcification spots are not accurately detected in MRI (in 1, 3 and 6-8), and that LRNC areas are better segmented when MRI is used (The relative volume is more accurate in 1-3, 5 and 6).

Table 3.4: Segmentation results when only MRI, CTA or distance features are used. Outlier rejection was performed before classifier training.*

Features (n)	Bias (% in result - % in GT)			Absolute error			Spearman (ρ)		
	C	F	LRNC	C	F	LRNC	C	F	LRNC
MRI (20)	-3.7±3.1	13.4±14.2	-9.7±12.6	3.7±3.1	15.8±11.2	12.8±9.2	-0.17	0.60	0.67
CTA (1)	7.9±5.6	27.5±18.5	-35.4±16.3	7.9±5.6	28.1±17.4	35.4±16.3	0.90	-0.03	0
MRI + d (23)	-3.6±3.3	9.5±13.7	-5.9±13.4	3.9±2.9	14.2±8.1	12.0±7.7	-0.05	0.81	0.79
CTA + d (4)	-0.4±1.1	6.8±16.3	-6.4±16.1	0.8±0.9	14.7±9.1	14.2±9.2	0.94	0.77	0.71
Distances (3)	-4.0±3.3	9.0±16.7	-5.0±16.7	4.1±3.1	15.5±10.3	14.4±9.1	-0.46	0.74	0.74
MRI + CTA (as in Table 3.3)	-0.2±1.4	7.6±13.0	-7.5±12.8	0.9±1.0	12.7±7.6	12.1±8.1	0.91	0.80	0.81

∗ d = distances, C = calcification, F = fibrous tissue, LRNC = lipid-rich necrotic core.

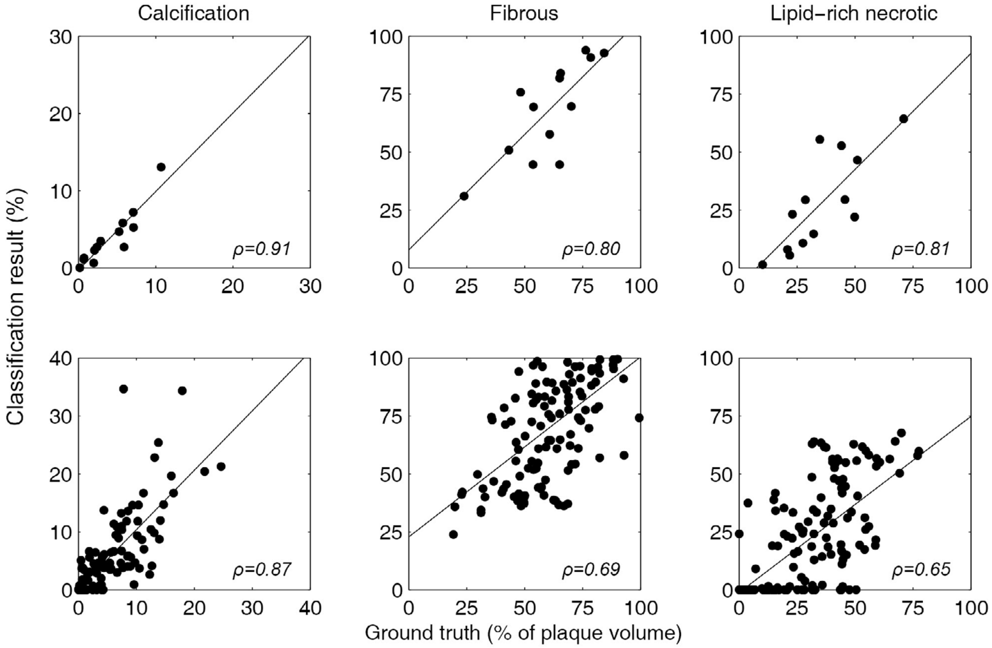


Figure 3.4: Correlation of plaque components in the ground truth and the classification result in 13 subjects. Here LDC and Gaussian outlier rejection were used. Top row: relative volumes per subject. Bottom row: relative volumes per slice.

3.3.4 Relevance of features

To give an indication of the most relevant features, the first five features selected by forward feature selection with LDC are provided in Table 3.5. It is clear that CTA intensity was the most important feature to segment calcification, and that distances and the combination of pre- and postcontrast images performed best to segment fibrous and lipid-rich necrotic tissue. Also first and second order derivatives showed to be relevant. This does not mean that these are the individually best performing features. For example, to differentiate fibrous tissue and LRNC, the individual features at position 2-6 (2DT1w blurred, lumen distance, PDw blurred, 3DT1w postcontrast GM and 2DT1w precontrast) that gave individually the highest LDC accuracy, were not found in the top 5 with forward feature selection.

Table 3.5: Feature selection. This table gives the first five features selected by forward selection using LDC accuracy as evaluation criterion, for the separation of each combination of two classes.

Calcification-Fibrous	Calcification-LRNC	Fibrous-LRNC
CTA intensity	CTA intensity	Distances multiplied
Distance to lumen	Distances multiplied	TOF Gradient Magnitude
3DT1w pre-contrast Laplacian	3DT1w pre-contrast Laplacian	3DT1w post-contrast Blurred
TOF Gradient Magnitude	Distance to lumen	3DT1w pre-contrast Blurred
PDw blurred	3DT1w post-contrast Blurred	3DT1w post-contrast Laplacian

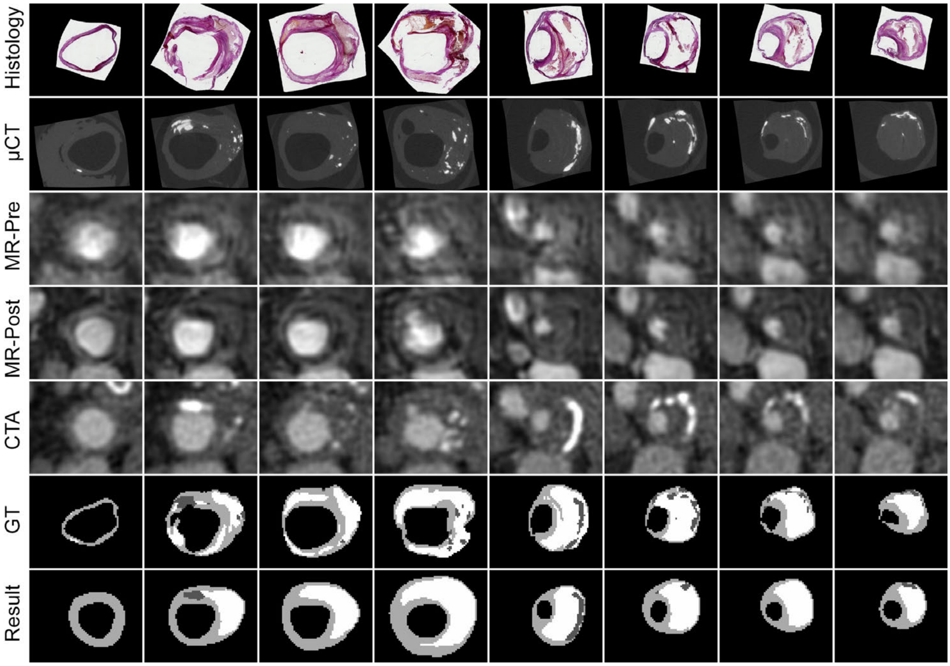


Figure 3.5: Segmentation results for one patient, using LDC and Gaussian outlier rejection. White = LRNC, light gray = fibrous tissue and dark gray = calcification.

3.4 Discussion

We segmented different components of atherosclerotic plaques using both *in vivo* MRI and CTA images, by training a voxelwise classifier on labels obtained from registered histology and μCT and taking into account the presence of misregistrations. Three different approaches (blurring and weighting by Dice overlap, weighting by contour distance, and Gaussian outlier rejection) showed that taking registration errors into account can improve component volume estimations in certain situations. In addition, we showed that combining MRI and CTA images results in better segmentations than when only MRI or only CTA is used.

For LDC no change was observed when sample weights were based on the local contour distances compared to the hard labels, but the bias for fibrous tissue and LRNC became smaller after blurring and weighting by the Dice overlap, and both the bias and absolute error decreased using Gaussian outlier detection. For SVM all three approaches improved the error and correlation for fibrous tissue and LRNC, but decreased classification accuracy for calcification. Weighting by the Dice overlap has a relatively large effect on the class priors (5% (Calcification), 56% (Fibrous) and 38% (LRNC) for hard labels, and 5%, 49% and 46% for weighting with Dice[24]), in contrast to weighting by contour distance and outlier rejection. This may have caused the change in bias of LRNC and fibrous tissue for LDC. Gaussian outlier rejection and LDC are both based on estimation of mean and covariance of Gaussian distributions, which may explain why the combination performs well. The advantage of outlier rejection was especially present in slices with a larger lipid content, which are the slices with more challenging composition that are more difficult to segment. For SVM, samples on the decision boundary determine the final segmentation result, which can be

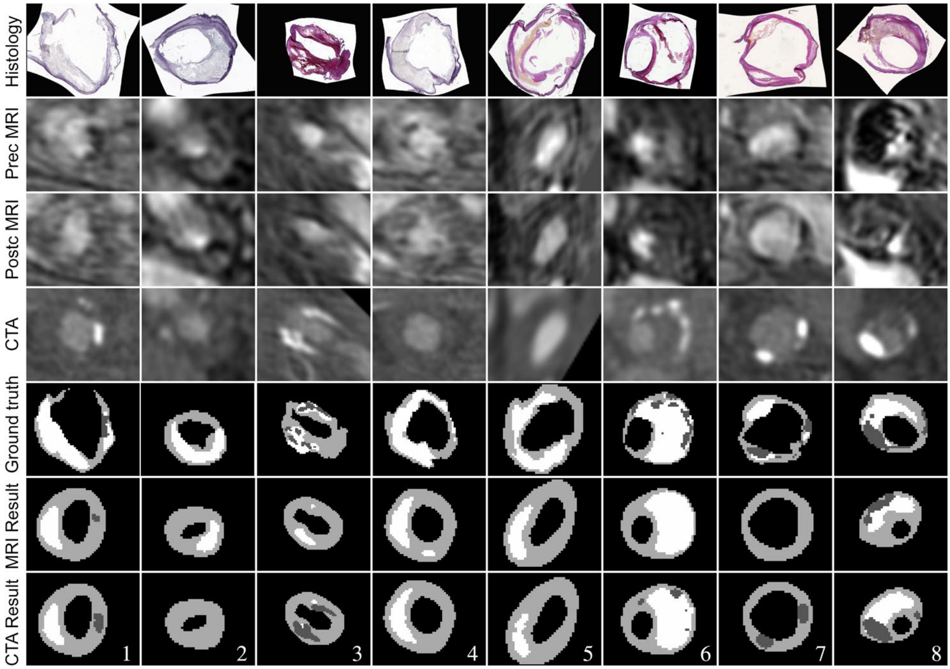


Figure 3.6: Segmentation results when only MRI, or only CTA, and distance features are used. Results are obtained including outlier rejection. White = LRNC, light gray = fibrous tissue and dark gray = calcification.

a reason why changing the weights based on local contour distance works better in this situation. LDC has been used previously for plaque component segmentation [219, Chapter 2], and performed better than SVM in our experiments. This indicates it is indeed a suitable classifier for this problem.

To compare our results with previous studies Table 3.6 can be used. Two previous studies automatically segmented plaque components in *in vivo* MRI and compared their results with histology [82, 119]. These show Pearson correlation values (R^2) of 0.83 for calcifications, 0.78 for necrotic tissue, 0.41 for loose matrix and 0.82 for fibrous tissue [119], and (R) 0.41 for calcifications, 0.75 for lipid, 0.61 for hemorrhage and 0.67 for fibrous tissue [82], compared to our values (R, see Table 3.6) of 0.92 (calcification), 0.78 (fibrous) and 0.79 (LRNC) using both MRI and CTA. In [219] plaque component segmentation results were compared with manual annotations of the *in vivo* data, which gave correlation values (R) of 0.88 for lipids, 0.80 for hemorrhage and for fibrous tissue and 0.10 for calcification. Although results are difficult to compare, our results are in a similar range. Hofman et al. [82] obtained less accurate classification results, but this study did not use any spatial information such as distance to the vessel wall. The method by Liu et al. [119] yielded high accuracies and is available within a commercial software package for plaque analysis [98]. This method is also based on voxel classification, and is followed by a level-set segmentation, resulting in more smoothly segmented regions. Spatial regularization is in our case achieved by using Gaussian features and distances and we found this leads to spatially coherent segmentations. The higher accuracies by [119] can be caused by the use of histology-guided manual contours. These are based both on histology and on known MRI intensities, which may

Table 3.6: Comparison to previous studies.

Study	Data	Evaluation	Results
Our study	13 subjects (144 slices) MRI and CTA Leave-one-subject-out cross-validation	Histology % of total volume per vessel	Calcium: ρ=0.91, R=0.92, ICC=0.92 Fibrous: ρ=0.78, R=0.78, ICC=0.76 LRNC: ρ=0.81, R=0.79, ICC=0.76
Liu et al., 2006 [119]	12 subjects (58 slices) MRI 14 subjects (84 slices) for training	Histology-guided manual contours Area (mm^2) per slice	Calcium: R^2=0.83 Fibrous: R^2=0.82 Loose matrix: R^2 = 0.41 Necrotic: R^2=0.78
Hofman et al., 2006 [82]	13 subjects (89 slices) MRI 12 subjects for training	Histology % of total volume per vessel	Calcium: R=0.44 Fibrous: R=0.69 Lipid: R=0.74 Hemorrhage: R=0.63
van 't Klooster et al., 2012 [219]	40 subjects (344 slices) MRI 20 subjects for training	Manual annotations Volume (mm^3) per vessel	Calcium: R=0.1, ICC=0.1 Fibrous: R=0.8, ICC=0.8 Lipid: R=0.88, ICC=0.65 hemorrhage: R=0.8, ICC=0.8
Wintermark et al., 2008 [237]	8 subjects (53 slices) CTA	Histology presence/absence 212 quadrants of 53 slices	Calcium: all correct Lipid: κ=0.495 (large areas: 0.796) Large hemorrhage: κ=0.712 Ulceration: κ=0.855
De Weert et al., 2006 [45]	14 subjects (41 slices) CTA	Histology % of area per slice	Calcium: R^2=0.74 Fibrous: R^2=0.76 Lipid: R^2=0.24

bias the segmentations towards intensity (gradients) seen in the MRI. A recent study [136] has shown that the LRNC appears smaller on MRI images than in histology, which is in correspondence with our segmentation results.

Similar to our experiments in which we only used MRI, previous studies that used MRI for segmentation, found low correlations for calcification [82, 219], except for the paper by Liu et al. [119]. Although calcification is difficult to detect in MRI, in Figure 3.6 dark spots can be seen at calcified locations. Reasons for poor detection in our study are low visibility in other slices, noise, dark-appearing artefacts that do not represent calcifications, and small misregistrations in the training data. The study of Liu et al. [119] obtained their ground truth by histology-guided manual annotation which eliminates the effect of misregistration. MRI sequences that are more specifically aimed at visualizing calcium, could also improve its detection [240].

Two previous studies compared automatic segmentation results in CTA with histology [45, 237], based on a fixed intensity threshold on the CTA to separate LRNC from fibrous tissue and fibrous tissue from calcification. Both these studies accurately segmented calcifications, but obtained lower accuracies for LRNC. Although de Weert et al. [45] found a significant difference between Hounsfield units for lipid (25±19) and fibrous tissue (88±18), the correlation for lipid volume is low (R^2=0.24, fibrous 0.76 and calcification 0.74). In mildly calcified (<10%) plaques the correlation for LRNC increased (R^2=0.77), which they relate to the blooming effect of calcification which may overshadow parts of soft plaque. Wintermark et al. [237] found overlapping Hounsfield units for lipid (32.6±20.0) and connective (fibrous) tissue (46.4±19.9). Concordance between CTA and histology in the detection of lipid tissue was therefore low (κ=0.495), but increased when only large lipid cores were included (κ=0.796). The difficulty they encountered to accurately segment lipid volumes was also observed in our study, which showed large errors when only CTA was used. In our experiments, the reasonable correlation with the ground truth for fibrous and lipid-rich necrotic tissue seems to be mostly based on the distances to the lumen and outer vessel wall, which on itself already yield plausible segmentations. Blooming artefacts in CTA have probably caused the overestimation of calcification when only CTA intensity was used. Blooming in the μCT did not affect the ground truth due to the relatively high resolution compared to the *in vivo* resolution.

Scanning patients with both MRI and CTA puts a higher burden on both patients and health-care costs and therefore the combination may not seem relevant in practice. However, in most cases patients that enter the hospital with symptomatic carotid artery disease are already scanned with CTA [23, 198]. MRI has no ionizing radiation, so is relatively safe for patients and can better determine the extent of non-calcified components of vulnerable or complicated plaques [159, 166]. Clinical studies showed that MRI and CTA have different advantages and that combining them may allow for more accurate decision on plaque vulnerability and treatment planning [74, 132]. Recently, clinical studies are emerging that perform both CTA and MRI imaging, showing the clinical possibility and relevance [190, 205, 207]. Ultrasound is another imaging modality that is feasible for imaging the carotid artery, due to its low cost and wide availability. Therefore ultrasound seems especially useful for screening. It can be used to study plaque vulnerability [109], but accurate quantification of plaque components is not possible due to the limited contrast [195]. For an accurate study of the vessel wall in high-risk patients MRI and CTA are recommended.

When the proposed segmentation framework is to be used for analysis of a new patient several steps need to be taken. These steps with their computation time on a desktop computer (2.26 GHz, 12.0 GB RAM) are: registration of MRI and CTA data (5 registrations, ~0.5 minute per registration) within a mask (~1-2 minutes for annotation on the CTA and T1w post-contrast MRI each), manual lumen and outer wall segmentation (~10 minutes, but this can be automated [209, 218]), inhomogeneity correction (~10-15 seconds per MR image), feature computation (~4-5 seconds), applying the classifier and obtaining segmentations (~1 second).

Our current results show a good Spearman rank correlation for the amount of LRNC, which is an indication of plaque vulnerability [245]. Whether this can reliably be used to select high-risk patients for treatment can, however, not be determined in this study. This has to be determined in a large group of patients that are followed for clinical events, followed by a clinical trial where the advantage of using plaque composition for treatment selection is evaluated. In our results the stroke patients had a LRNC of 44% (range 10-71), and the patients who had a TIA had a LRNC of 27% (22-46). The amount of calcification was 3% (0-6) after stroke, and 7% (1-11) after TIA. The automated results yielded similar results: a LRNC of 46% (1-64) and 23% (5-29), and calcifications of 3% (0-6) and 5% (1-13) for stroke and TIA patients respectively. This corresponds to the idea that a LRNC characterizes more vulnerable plaques and calcifications characterize more stable plaques [168, 188]. The asymptomatic patient in this study had a LRNC of 20% and 1% of calcification.

This work has several limitations. There was a considerable time interval between the MRI and CTA scan (38±26 days), but we do not expect noticeable changes in plaque composition and volume to occur in this period. Previous studies did not find changes in wall and component volumes or presence in a mean of 12-15 days [115, 210] or a year [105].

For this study no histology sections were stained to specifically detect intraplaque hemorrhage (IPH), and hemorrhage was therefore not included as a separate component. In addition, the used T1w MRI sequences are not the most suitable for imaging of IPH [141]. Still, based on hyperintensity in the precontrast 2D- and 3D-T1w MR images, we found a suggestion of IPH presence in 5 vessels (4 with stroke, 1 with TIA). For all these vessels this area was included in the LRNC segmentation. The error for LRNC was on average not larger in those 5 vessels than in the other 8 for the experiments including MRI and/or CTA and distance features. When only CTA intensity was considered no LRNC was segmented, both when all patients were included and when only the patients without possible IPH presence were included in training and evaluation. This suggests that IPH presence is not causing large errors, nor explains the poor performance of CTA in this study. Using the presented approach, hemorrhage can be easily added separately to the framework when a ground truth is available, and its inclusion would be highly valuable in future studies [188]. We also did not differentiate between fibrous and loose connective tissue as was done in [119], as these are difficult to distinguish in histology and both are stable plaque components. Before histology sectioning the specimens were decalcified, and during sectioning areas of lipid tissue may be disrupted, which could cause mixing up the two tissues. However, in the empty regions in the histology sections we could distinguish calcium from lipid by using the μCT as a reference.

In this work we evaluated our results based on relative plaque component volumes. Ideally, the classification accuracy would be evaluated in a voxelwise manner. We could not do this because no accurate voxelwise correspondence could be established between *in vivo* data and histology, even with substantial manual interaction. As a surrogate measure, we chose relative plaque component volumes as they provide a clinically relevant biomarker for prediction of plaque vulnerability [104, 186, 239, 245]. Additionally, there were only 13 subjects included in this study. However, due to the challenges involved in plaque excision, sectioning and staining, it was not possible to add more vessels.

Lastly, in this work we use manual annotations of the vessel wall in *in vivo* images. Several automated vessel wall segmentation methods are available [14, 98, 209, 218]. The combination of automated wall segmentation with component segmentation would yield a highly automated plaque analysis tool. As long as an automatic wall segmentation is accurate, and possibly manually adjusted in case of errors, we expect using an automatic segmentation method has no influence on the results described in this paper.

3.5 Conclusion

The volume of atherosclerotic plaque components can be well estimated using a classifier trained on histology. Different approaches to account for mismatch between the imaging data and the ground truth from histology can slightly improve segmentation. While MRI can better differentiate between fibrous and lipid-rich necrotic tissue, and CTA can better segment calcification, the combination leads to good results for all three components. This can facilitate the use of quantitative plaque composition in large clinical studies and possibly future patient risk assessment.

Acknowledgements
We thank Kim van Gaalen for her technical assistance.

Chapter 4

Automated segmentation of atherosclerotic histology based on pattern classification

This chapter has been published as:
A. van Engelen, W.J. Niessen, S. Klein, H.C. Groen, K. van Gaalen, H.J.M. Verhagen, J.J. Wentzel, A. van der Lugt and M. de Bruijne, Automated segmentation of atherosclerotic histology based on pattern classification, *Journal of Pathology Informatics*, 2013

◆

Abstract

Background: Histology sections provide accurate information on atherosclerotic plaque composition, and are used in various applications. To our knowledge, no automated systems for plaque component segmentation in histology sections currently exist.

Materials and Methods: We perform pixelwise classification of fibrous, lipid, and necrotic tissue in Elastica Von Gieson-stained histology sections, using features based on color channel intensity and local image texture and structure. We compare an approach where we train on independent data to an approach where we train on one or two sections per specimen in order to segment the remaining sections. We evaluate the results on segmentation accuracy in histology, and we use the obtained histology segmentations to train plaque component classification methods in *ex vivo* Magnetic resonance imaging (MRI) and *in vivo* MRI and computed tomography (CT).

Results: In leave-one-specimen-out experiments on 176 histology slices of 13 plaques, a pixelwise accuracy of 75.7±6.8% was obtained. This increased to 77.6±6.5% when two manually annotated slices of the specimen to be segmented were used for training. Rank correlations of relative component volumes with manually annotated volumes were high in this situation (ρ=0.82-0.98). Using the obtained histology segmentations to train plaque component classification methods in *ex vivo* MRI and *in vivo* MRI and CT resulted in similar image segmentations for training on the automated histology segmentations as for training on a fully manual ground truth. The size of the lipid-rich necrotic core was significantly smaller when training on fully automated histology segmentations than when manually annotated histology sections were used. This difference was reduced and not statistically significant when one or two slices per section were manually annotated for histology segmentation.

Conclusions: Good histology segmentations can be obtained by automated segmentation, which show good correlations with ground truth volumes. In addition, these can be used to develop segmentation methods in other imaging modalities. Accuracy increases when one or two sections of the same specimen are used for training, which requires a limited amount of user interaction in practice.

4.1 Introduction

Stained histological sections contain highly detailed information on tissue composition. Because of a resolution in the order of micrometers and the availability of a variety of specific stains, human observers can accurately indicate which cells or tissue types are present at each location. One of the applications for which histology is used, is the study of atherosclerotic plaques. It has been shown that the composition of such plaques is related to plaque vulnerability, indicating its chance to rupture and possibly causing cerebrovascular events [167]. Histology sections are segmented into different tissue components to improve the understanding of risk factors of stroke [167], for biomechanical modeling of stress distributions in the plaque [6], and to develop and evaluate plaque component segmentation methods on *ex vivo* and *in vivo* MRI [38, 119, Chapter 2, Chapter 3].

Manual annotation of histology sections is a time-consuming task. Several studies have automatically segmented histology sections, mainly for oncological applications [71]. To the best of our knowledge, no studies on automated segmentation of atherosclerotic plaque components in histology have been published. Image analysis in histology often consists of segmenting typical structures such as lymphocytes, cell nuclei and glands, and/or indicating a presence or severity of disease [71]. Our approach is different from this in the sense that we aim to divide the complete section into different areas. These segmentations can then be used to determine the stage of atherosclerotic disease, or to train plaque component segmentation methods in registered imaging data [Chapter 2, Chapter 3].

In this paper we (semi-)automatically segment histology sections of carotid atherosclerotic plaques into fibrous (stable plaque component), necrotic and lipid (vulnerable components) regions using pixel-wise classification. Although all evaluated sections are stained with the same protocol, variations in appearance occur between sections from different specimens. Therefore we compare an approach that uses training data from different specimens to training on one or two slices of the specimen to segment, which in practice requires a limited amount of user interaction per specimen. We also evaluate whether these segmentations can reliably be used to train plaque component segmentation methods in *ex vivo* MRI and *in vivo* MRI and computed tomography-angiography (CTA), using datasets in which spatial correspondence between these imaging data and histology has been established.

4.2 Methods

4.2.1 Data

For 13 patients who were scheduled for carotid endarterectomy, *in vivo* MRI (pre- and post-contrast 3DT1-weighted, precontrast T1-weighted, proton density weighted (PDw) and Time of Flight (TOF)) and CTA scans were acquired. After surgery, the excised specimens were scanned with MRI (3DT1w, PDw, T2w) and μCT, and afterwards decalcified and embedded in paraffin for histology processing. At every 1-mm interval, axial slices of $5\,\mu$m thickness were taken. These sections were stained with Elastica Von Gieson (EvG, Merck, Germany). This stain colors nuclei brown to black, collagen and elastin pink to red, and muscle and cytoplasm yellow. Lipid regions are characterized by the typically shaped cholesterol esters and the absence of connective tissue. Necrotic tissue contains no cell nuclei and is usually less dense than fibrous tissue. Examples, which also depict the variability between sections, are shown in Figure 4.1.

On the digitized histology slices (resolution 1.8×1.8 μm) contours for the lumen, outer vessel wall, necrotic tissue, and lipid regions were manually drawn. Areas within the vessel wall that were not annotated as necrotic or lipid regions were labeled as fibrous tissue. For registration and evaluation purposes lumen and outer wall contours were drawn on the *ex vivo* MR images and the *in vivo* aligned MR and CTA images. Sections with low image quality due to histology disruption and sections which contained almost no tissue were excluded. In total, 176 histology sections were used for analysis (13.5±4.9, range 5-21, per specimen).

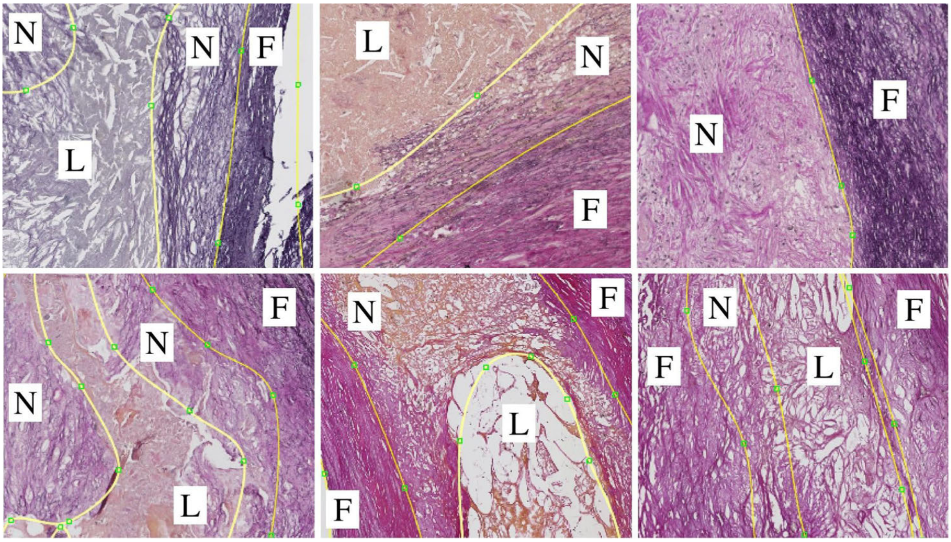


Figure 4.1: Examples of histology sections and manual segmentations that show the variation in our dataset. F = Fibrous, N = Necrotic, L = Lipid.

4.2.2 Histology Pattern Classification

Image Features

Color, texture and tissue structure are important features to distinguish plaque components (Figure 4.1), and therefore features of these types were used. Color: we used intensity of the 3 red, green, blue (RGB) color channels before and after Gaussian smoothing (σ=0.1 mm), where in general fibrous tissue is dark purple, necrotic tissue lighter purple and lipid more pink or yellow. Texture: the mean gradient magnitude at σ=1 pixel and the standard deviation, both of the green channel in square windows of 10×10, 20×20 and 30×30 pixels, were expected to differentiate tissues by tissue heterogeneity and the strength and amount of edges. The structure tensor [233] was calculated (inner scale σ=1pixel=1.8 μm, outer scale σ=72 μm) to differentiate between highly structured tissues (parallel fibrous bands) and less ordered tissues (lipid, and to a lesser extent necrotic tissue). We used the first and second eigenvalue ($|\lambda_1| > |\lambda_2|$), coherence ($\sqrt{\lambda_1{}^2 - \lambda_2{}^2}$), and the determinant.

The original histology slices have large dimensions and are, therefore, time-consuming to process. The color and texture features were calculated from downsampled images that still had a resolution of 0.015×0.015 mm. The features of the structure tensor were calculated from the original images, and were subsequently downsampled as well to perform training and classification.

Segmentation

Pixel classification was performed within the manual contours of the vessel wall, with linear discriminant analysis (LDA) for the three components fibrous, lipid, and necrotic tissue. Calcified tissue was not included as the specimens were decalcified and calcifications were not available in the manual annotations. Two approaches were evaluated: 1) training a classifier on an independent dataset and 2) training a classifier on one or two sections of the specimen to be segmented (details in Section 4.2.3). Whereas, the second approach requires the annotation of one or two slices for each new specimen, this approach is expected to better account for differences in appearance between specimens. Therefore, of each specimen two histology sections were selected: The first was thought to be most representative for the other sections, and the second to form the most representative combination together with the first selected section. Representative sections contained all three tissue components, and the tissue of these components was similar in appearance to the other sections of that specimen.

4.2.3 Evaluation

Histology segmentation

Classification accuracy by LDA was compared between: 1) Leave-one-specimen-out cross-validation, where 1% of the pixels from the twelve other specimens were used for training, 2) training on all pixels from the first selected section of the specimen to be segmented, and 3) training on both selected sections. Evaluation was performed on the sections that were never used for training, so on the same sections in every case (number of slices per specimen minus two). The average pixel-wise accuracy for 13 specimens, overall sensitivity for all three components for all specimens together, and correlation of relative component volumes with those obtained from manual segmentations were calculated. A Friedman test was done to test the accuracy for statistical differences.

Ex vivo MRI

To evaluate the validity of the obtained histology segmentations for training of *ex vivo* MRI plaque segmentation methods, we use the approach proposed in Chapter 2. This method uses registered three-dimensional (3D) histology to train a voxel classification method. The lipid and necrotic regions from histology were combined to form lipid-rich necrotic cores (LRNC), which is often done in MRI studies [38, 119, Chapter 2]. Areas for calcification were obtained by thresholding μCT scans, and subtracted from the (automatically determined) fibrous and LRNC segmentations, resulting in three components for classification. Results were compared for training on four different ground truth histology segmentations: 1) the manual histology segmentations, 2) the leave-one-out automated segmentations, 3) the manual histology segmentation for the first selected section per specimen, and the automatic segmentations after training on this section for the other sections, and 4) the manual histology segmentations of both selected sections, and the automatic segmentations after training on these sections for the remaining sections.

Histology, μCT, and *ex vivo* MRI were registered using a semi-automatic 3D registration approach [68, Chapter 2]. The same feature set as previously was used [Chapter 2], consisting of the available MR images (normalized), first and second order derivatives, and distances to the vessel wall. LDA was evaluated in a leave-one-specimen-out manner, training on 1%

Table 4.1: Results for histology classification. F=Fibrous, L=Lipid, N=Necrotic.

Training method	Accuracy	Sens. F	Sens. L	Sens. N	Sens. L+N	ρ (F,L,N)
Leave-one-out	75.7±6.8%	90.3%	76.4%	33.2%	72.9%	0.88, 0.93, 0.74
First selected slice	76.7±6.5%	86.8%	85.7%	42.4%	77.2%	0.98, 0.97, 0.78
Two selected slices	77.6±6.3%	87.7%	86.6%	45.3%	78.8%	0.98, 0.98, 0.82

of voxels of the other twelve specimens, for each specimen.

For evaluation, voxelwise accuracy and the relative plaque component volumes were determined per specimen (all slices combined) with respect to the manual histology segmentations and thresholded μCT. A Friedman test was done to test the voxelwise accuracy and volume errors for statistical differences.

In vivo MRI and CTA

Combined MRI and CTA images were segmented as proposed in Chapter 3, using the four different histology segmentations for training, as explained in the previous subsection. Image features were the same as previously [Chapter 3] and similar to the features used in the *ex vivo* image segmentation. Registration of all imaging modalities and histology was carried out as described previously [68, Chapter 3]. As the final registration between histology and *in vivo* data is less accurate than between histology and *ex vivo* data, evaluation was only done on Spearman rank correlation of relative volumes and on the error of the relative volume of each component in the result compared to the manual ground truth, per specimen (all slices combined per specimen).

4.3 Results

4.3.1 Histology Segmentation

Results are shown in Table 4.1 and Figure 4.2. With training on independent data an accuracy of 75.7±6.8% was obtained. Spearman rank correlations were high for fibrous (0.88) and lipid (0.93), and good for necrotic tissue (0.74). Using training data from the same specimen slightly increased both accuracy (77.6±6.3%) and correlation (ρ=0.98 for fibrous and lipid and 0.82 for necrotic tissue), although not statistically significant. The main improvement was an increased sensitivity for lipid and necrotic tissue. As for MRI lipid and necrotic tissues are often combined into one vulnerable component, the combined sensitivity is also given. Specificity for fibrous tissue is then equal to this sensitivity, and vice versa.

4.3.2 *Ex vivo* MRI

Results are shown in Table 4.2 and Figure 4.3. The results were very similar for training on either manual or automated histology segmentations, and the accuracies and correlations showed no statistically significant differences. The errors in the estimated volumes of the three components were only significantly larger than training on manual histology segmentations when training on leave-one-specimen-out segmented histology, and for calcification. The difference for calcification was so small that it is clinically probably not relevant.

Table 4.2: Results on *ex vivo* MRI. C=Calcification, F=Fibrous, LRNC=Lipid-rich necrotic core, ρ=Spearman's rho, Error = Absolute (% in result - % in ground truth). *Significant difference with training on manual histology segmentations (P<0.05).

Histology segm.	Accuracy	ρ (C, F, LRNC)	Error (C, F, LRNC)
Manual	72.7±6.8%	0.86, 0.76, 0.76	2.4±2.4, 11.2±8.4, 13.0±9.4
Leave-one-out	71.3±7.8%	0.86, 0.79, 0.75	2.6±2.5*, 15.2±9.8*, 17.9±10.0*
Trained on 1 slice	71.8±7.1%	0.85, 0.83, 0.78	2.7±2.5*, 11.6±8.8, 14.0±9.5
Trained on 2 slices	72.3±6.7%	0.85, 0.80, 0.78	2.8±2.6*, 10.6±7.8, 12.7±9.0

4.3.3 *In vivo* MRI and CTA

Results are shown in Table 4.3 and Figure 4.4. Spearman rank correlation values of relative plaque component volumes for the 13 specimens are similar for training on the automated and manual histology segmentations and even show slight improvements. The absolute error in volume estimation of fibrous and lipid-rich necrotic tissue is somewhat larger when the classifier is trained using the automated histology segmentations. This is not significantly different when the histology is segmented by training on data from one or two slices per specimen.

4.4 Discussion and Conclusion

We performed automated segmentation of histology sections of atherosclerotic plaques, and subsequently used these as a ground truth to train segmentation algorithms in *ex vivo* and *in vivo* images.

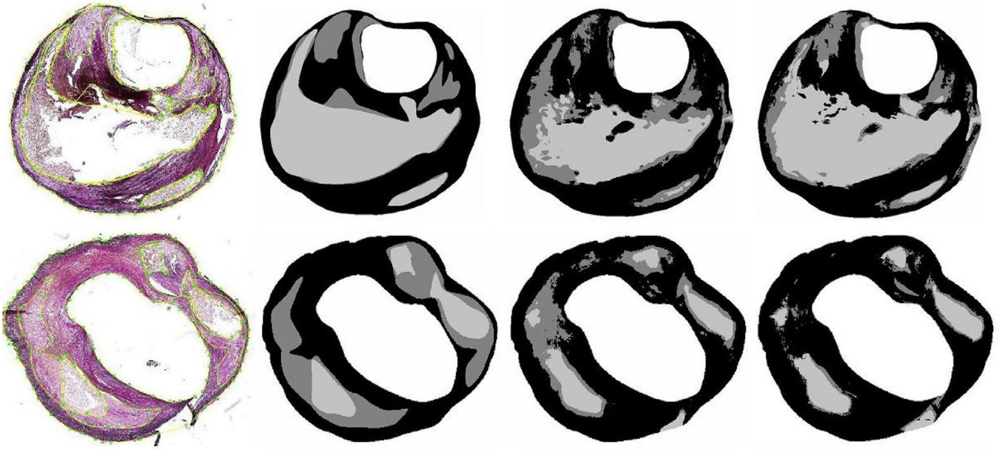


Figure 4.2: Examples of histology segmentation. From left to right: Histology section, manually annotated ground truth, result by leave-one-out training, and the result for training on two other sections of the same specimen. In the top row, the accuracy is 73.3% with leave-one-out training, and 82.2% when trained on two other slices of that specimen. All tissue types are well-detected. In the bottom row, the accuracies are 66.7% and 67.8%. The sensitivity for necrotic tissue is lower. Black = fibrous, dark gray = necrotic, and light gray = lipid.

Table 4.3: Results on *in vivo* MRI. C=Calcification, F=Fibrous, LRNC=Lipid-rich necrotic core, ρ=Spearman's rho, Error = Absolute (% in result - % in ground truth). *Significant difference with training on manual histology segmentations (P<0.05).

Histology segmentation	ρ (C, F, LRNC)	Error (C, F, LRNC)
Manual	0.87, 0.84, 0.81	1.0±1.2, 14.4±7.3, 14.0±7.2
Leave-one-out	0.89, 0.89, 0.90	1.0±1.1, 19.8±10.2*, 19.3±10.1*
Trained on 1 slice	0.87, 0.87, 0.84	1.0±1.2, 16.9±8.4, 16.5±8.2
Trained on 2 slices	0.87, 0.87, 0.87	1.0±1.2, 15.6±7.8, 15.2±7.7

The good accuracy and high Spearman rank correlations of histology segmentation indicate that, although multiple factors are relevant, the automated histology segmentations may be used as a measure in studying plaque vulnerability. The accuracy slightly improved by training on one or two slices of the specimen to be segmented instead of training on data from other specimens, however, differences were not significant.

When these automated histology segmentations were used to train *ex vivo* and *in vivo* plaque component segmentation methods, very similar accuracies and correlations were found as when the manual histology segmentations were used for training. This suggests a similar performance for grading of plaque vulnerability. Fibrous tissue was a little more overestimated and LRNC more underestimated when the automated segmentations were used. This is mainly because in the histology sections the difference between fibrous and necrotic tissue is difficult to detect consistently, which results in the classification of necrotic tissue as fibrous. MRI segmentation is slightly better when the automated histology sections for training were obtained by training on one or two annotated sections per

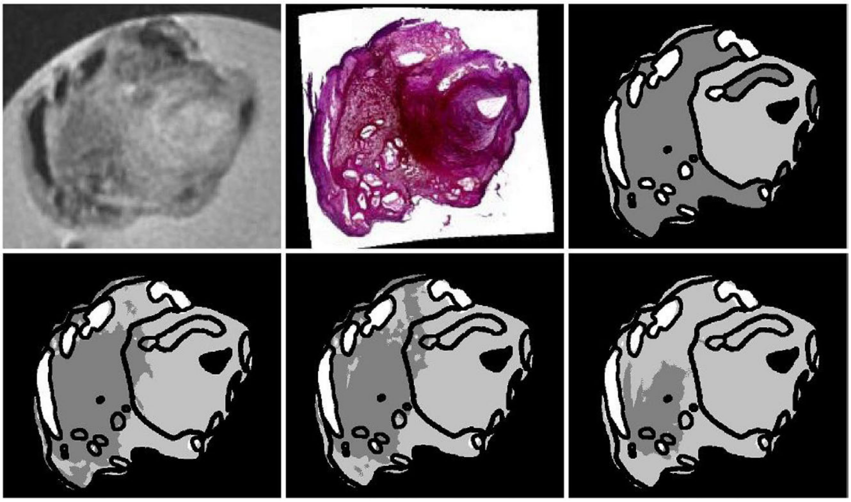


Figure 4.3: Example of *ex vivo* MRI segmentation. Top row from left to right: PDw MRI, histology section and the ground truth. Bottom row: Similar results are obtained by training on manually segmented histology (left figure, accuracy 74.6%) or histology segmentations trained on two slices of that specimen (middle figure, 75.1%). The histology sections segmented using leave-one-out classification result in a reduced accuracy (right figure, 61.4%). Light gray = fibrous, dark gray = LRNC, and white = calcification.

specimen. Both better histology segmentations, and the use of the manually annotated sections in the ground truth for MRI classifier training contribute to this. Annotation of two slices still would reduce the number of histology sections that need complete manual segmentation by 85% in the used dataset.

Segmentation of plaques into hard classes may be difficult at specific locations, as gradual changes from one component to another occur. This can be taken into account in future research by including a mixed class or by obtaining posterior probabilities instead of hard segmentations. In addition, empty regions may be present in histology. Lipid tissue is very soft and therefore may be disrupted during slicing and observers annotate empty regions usually as lipid. In relation to interobserver reproducibility of histology segmentation our results are in between previously reported values. The same annotations for 92% of all pixels by two observers has been reported [38], but moderate agreement for observers indicating which components or characteristics are present has also been found [123] (κ: 0.44-0.68 for different components).

This paper uses manual lumen and outer wall contours in all histology sections and is therefore not completely automatic. However, as the specimen can clearly be distinguished from the background, we expect automatic segmentation to be relatively easy.

In conclusion, good histology segmentations can be obtained using automated pixel classification. Automatic histology segmentation by using independent training data seems sufficient to indicate plaque vulnerability in clinical studies. These segmentations can also be used as training data to design automatic techniques for segmentation of plaque components in *ex vivo* MRI and *in vivo* MRI and CTA data. By using several manually annotated sections for each specimen the segmentations can be improved.

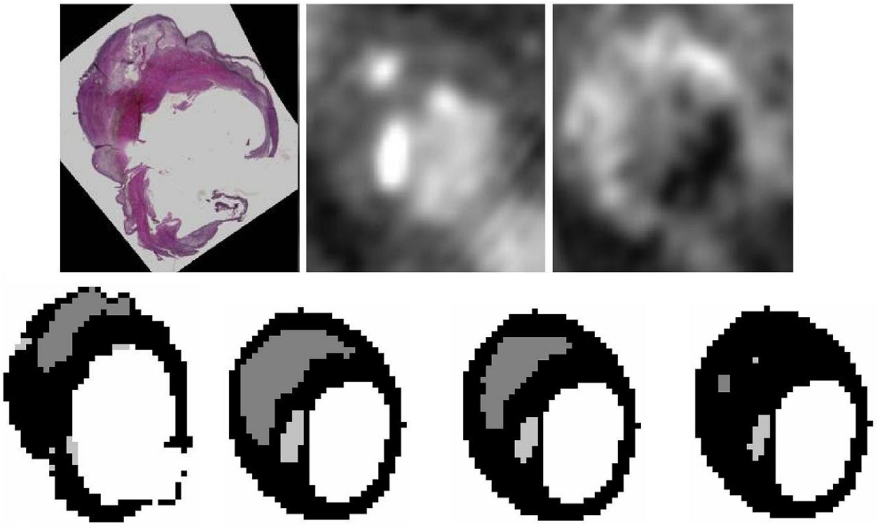


Figure 4.4: Example of *in vivo* MRI/CTA segmentation. Top row from left to right: Histology section, CT-angiography and PDw MRI. Bottom row: The relative size of the LRNC is 15.0% in the ground truth (left figure), 27.4% using the manual histology segmentations (second figure), 17.7% for training on two slices of the specimen (third figure), and 1.0% for leave-one-out histology segmentation (right figure). Black = fibrous, dark gray = LRNC, and light gray = calcification.

Chapter 5

Multi-center MRI plaque-component segmentation using feature normalization and transfer learning

This chapter is based on:
A. van Engelen, A.C. van Dijk, M.T.B. Truijman, R. van ’t Klooster, A. van Opbroek, A. van der Lugt, W.J. Niessen, M.E. Kooi and M. de Bruijne, Multi-center MRI plaque-component segmentation using feature normalization and transfer learning, *Submitted*

* * *

Abstract
Automated carotid atherosclerotic plaque component segmentation in MRI is important to enable large studies on plaque vulnerability, and will be key for incorporating it as an imaging biomarker in clinical practice. These methods facilitate the quantification of plaque components, and especially supervised techniques which learn from examples have shown good performance. However, a disadvantage of many supervised methods is their reduced performance on data acquired on different scanners with possibly different settings. Ways to overcome the need to acquire many manual annotations for each new dataset can stimulate widespread implementation of supervised methods. In this paper we segment carotid plaque components of clinical interest (fibrous tissue, lipid tissue, calcification and intraplaque hemorrhage) in a multi-center MRI study. We perform voxelwise tissue classification by same-center training, and compare results with two approaches that either use no or few annotated samples from the same-center data in addition to an annotated set of different-center data. First, we evaluate a non-linear feature normalization approach, and secondly two Transfer-Learning algorithms are evaluated. Results showed that feature normalization can improve the ability to segment new data when only different-center training data is available. Transfer learning can achieve this using a limited number of same-center training data. Best results were obtained for a combination of feature normalization and transfer learning. We conclude that both extensive feature normalization and transfer learning can be valuable for the development of supervised methods that perform well on different types of datasets, without the need to obtain a large set of labels for each dataset.

5.1 Introduction

As the underlying cause of cardiovascular and cerebrovascular ischemia, atherosclerosis is one of the largest causes of mortality and morbidity worldwide [180, 235]. One mechanism by which atherosclerosis can lead to cerebral ischemia is rupture of carotid-artery plaques [65, 146]. The general consensus is that rupture-prone vulnerable plaques are characterized by a thin or ruptured fibrous cap, a large lipid-rich necrotic core (LRNC), presence of intraplaque hemorrhage (IPH) and active inflammation [131, 133, 160]. Additionally, a larger degree of calcification (CA) is related to a lower risk of plaque rupture [104]. In current clinical practice these factors are largely ignored and the decision to perform surgical treatment is still based on the degree of vessel narrowing as determined by non-invasive imaging [26, 143]. It has been hypothesized that determining plaque composition can help in assessing the risk of rupture and thereby improve the selection of patients for intervention [26, 152, 228].

Due to its superior soft-tissue contrast, magnetic resonance imaging (MRI) is the preferred imaging modality to visualize the different tissues in the atherosclerotic carotid vessel wall [125, 166]. The appearance of plaque tissues in different MR image sequences has been well established with respect to histology [77, 158, 243], and plaque components as measured from MRI have been related to future cerebrovascular events [70, 160, 188].

Automated segmentation of plaque components is essential for a quantitative analysis of plaque composition in large research studies and would greatly facilitate possible implementation of carotid MR imaging in daily clinical practice. Several methods have been proposed [82, 119, 219, Chapter 3]. These are all supervised methods that obtained a training set with either labels from histology, or from manual annotations. All perform voxel classification using MRI intensities, intensity gradients and wall distances as features. Liu et al. [119] used Parzen window estimation in a naive-Bayesian network and Hofman et al. [82] compared different approaches of which a quadratic Bayesian classifier performed best, while others [219, Chapter 3] used a linear discriminant classifier. These methods obtained reasonable to good results, varying between components. However, a limitation of such supervised methods is that they are specifically optimized for data acquired on one type of scanner using one specific imaging protocol. They assume training and target data follow the same distribution, which requires a large training set with annotations for each dataset with different properties. This raises problems when training and target data are different, for example when the MR sequence protocol changes, a scanner is replaced, or in multi-center studies, since images then typically have different contrast characteristics. Ideally, one would like an algorithm that is developed on one dataset to be applicable to datasets with different properties, without needing manual contours from these other datasets, which are labor-intensive to acquire.

Some approaches to overcome this problem have been investigated. Fischl et al. [56] incorporated physics of the MRI acquisition into the segmentation algorithm. Theoretically, with knowledge of intrinsic tissue properties (T1 and T2 relaxation time, and tissue proton density), tissue appearance can be modelled given any MR settings. However, these intrinsic properties are often unknown. Another approach involves image normalization. Normalization by matching the mean and variance of the image intensities or by matching two points from the intensity histogram is commonly used, mainly to handle differences between subjects scanned with the same parameters on the same scanner. More elaborate methods have been used to handle differences between scanners or protocols. For example matching more points from the MRI intensity histogram [138] resulted in better performance of segmentation with training and test data from different sources [247]. Artan et al. [15] applied a classifier trained on data from one device to data from a different device using iterative classification and intensity rescaling of the target data. For chest radiographs and chest computed tomography (CT), normalization of scans acquired with different settings by splitting and weighting different frequency bands, has shown to improve segmentation performance [61, 147]. Instead of normalizing the images, which may be dependent on the chosen region of interest or other structures present in the images, we present a way of feature normalization that is able to handle non-linear scaling of feature spaces from different sources.

Transfer learning [142] is an approach that is still relatively new to the field of medical image analysis. Transfer learning comprises machine-learning methods that handle differences in distributions, labeling functions, and/or features between training and test data. Transfer-learning methods have the potential that supervised segmentation methods can work with no or much less labeled training data with the same properties as the data to be analyzed (target data), by using more previously acquired and labeled data that originates from a different scanner, or was acquired using different settings (source data). For example, Wu et al. [238] used a weighted Support Vector Machine (SVM) and a weighted k-Nearest-Neighbour (kNN) in which source and target samples are weighted differently. Ablavsky et al. [1] present an approach for the segmentation of microscopy images, where a classifier trained on data with different properties than the target data was used to regularize a SVM classifier trained on a small set of labeled target data. For brain tissue segmentation, Van Opbroek et al. [216] proposed a reweighting SVM where the weight of misclassified source samples was iteratively reduced in order to reduce the influence of source samples that contradict the rest of the data, while maintaining informative source samples. These examples have shown the advantage of transfer-learning methodologies when little training data from the target data type is available.

In this study we aim to develop methods for plaque-component segmentation on multi-center MRI data that has been acquired on MRI scanners from different vendors, and with significant differences in MRI pulse sequence implementation. We implement different feature normalization and transfer-learning strategies, and evaluate I) the performance of voxel classification when training and test data come from the same center as well as from different centers, II) the performance of transfer learning, while using a small number of labeled samples from the target data and a large set of annotated source data, and III) the effect of extensive feature normalization.

5.2 Methods

5.2.1 General segmentation methodology

All experiments were performed using linear discriminant classification (LDC) of all voxels within the vessel wall, as was successfully applied previously [219, Chapter 2, Chapter 3]. The class conditional density ρ_k for class k is defined as follows by LDC [76]:

$$\rho_k(\mathbf{x}) = \mathbf{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k - \frac{1}{2} \boldsymbol{\mu}_k^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k + \log \pi_k \tag{5.1}$$

Here $\mathbf{x}$ is the feature vector to classify, $\boldsymbol{\Sigma}$ the covariance matrix that is pooled over the classes, $\boldsymbol{\mu}_k$ are the class means, and π_k the class prior probabilities. For classification, each sample is assigned to the class with the highest class density ρ_k. To obtain posterior probabilities the k conditional densities are normalized to sum to one.

In our experiments we used four classes: fibrous tissue (FT), LRNC, CA and IPH. Similar to Chapter 3, normalized image intensities of several MRI sequences, the images blurred with a Gaussian filter (σ=0.3 mm (=1 voxel in the data from center 1)), the gradient magnitude and Laplacian at that same scale, the Euclidean distances to the lumen and outer vessel wall (mm), and the product of these two distances were used as features. This resulted in a total of 23 features. All classifier training and evaluation was performed using Matlab (Release 2011b, The MathWorks, Inc., Natick, Massachusetts) and the prtools toolbox [50].

5.2.2 Adaptive histogram binning

A feature normalization step was performed to account for non-linear intensity scaling differences between imaging protocols. For each feature we took all voxels from the vessel wall from all patients. An adaptive histogram binning using piece-wise linear rescaling was applied to each feature independently, separately for each imaging protocol. All samples (voxels) were evenly distributed over 100 bins from 0 to 1000, such that each bin contained 1% of the samples. Here we assumed that the data from each center had a similar distribution with respect to patient characteristics, and hence a similar fraction of FT, CA, LRNC and IPH voxels. It also assumes that for each feature the ordering of tissue components in the imaging protocols is the same, but the contrast between tissues may vary. This procedure also affects the distribution of samples in the feature space as regions in the original histograms with high density are stretched out over more bins.

5.2.3 Transfer learning

We propose two forms of transfer learning for use with LDC, inspired by the sample weighting transfer-learning approaches by Wu et al. [238] and Van Opbroek et al. [216]. For both approaches the training data was composed of a large labeled dataset acquired in a different center than the data we aim to segment (called the different-center data), and a small number (n) of labeled samples from data acquired in the center for which we aim to segment the data (called same-center data). The labeled samples can for example be obtained by manually indicating a few locations of the different tissue types, or by segmenting a number of slices manually. We propose *weighted-LDC* and *reweighted-LDC*, which both use LDC as provided in equation 5.1. By giving individual samples a certain weight, the estimated class means $\boldsymbol{\mu}_k$ and the pooled covariance matrix $\boldsymbol{\Sigma}$ are affected:

$$\boldsymbol{\mu}_k = \frac{\sum\limits_{i:y_i=k} w_i \mathbf{x}_i}{\sum\limits_{i:y_i=k} w_i} \qquad \Sigma = \sum_{k=1}^{4} \pi_k \frac{\sum\limits_{i:y_i=k} w_i (\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T}{\sum\limits_{i:y_i=k} w_i} \tag{5.2}$$

with y_i the label y, and w_i the weight of sample i. For both $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}$ the contribution of each sample is determined by its sample weight. The classes contribute differently to the pooled covariance based on their prior probability π_k.

With *weighted-LDC* the total sum of the sample weights of the different-center data (ΣW_{dc}) and the sum of the sample weights of the same-center data (ΣW_{sc}) were set, while all samples from the same center had the same weight. In our experiments we kept ΣW_{sd} fixed at 1, and varied (ΣW_{sc}). Weighting the data from two centers differently determines the balance between the contribution of different-center and same-center data to the classifier. For *reweighted-LDC* we first apply weighted-LDC according to equations 5.1 and 5.2. After this step, for each different-data sample ρ_k was determined for its label k, and used as new sample weight. Per class the weights of all samples were linearly rescaled such that their sum equalled the initial total weight of that class. So, also the ratio between ΣW_{sc} and ΣW_{sd} remained the same as for weighted-LDC, only the weights of the different-center data varied between samples:

$$w_i = \frac{\rho_{y_i}(\mathbf{x}_i) \cdot \pi_{y_i}}{\sum\limits_{i:y_i=k} \rho_{y_i}} \tag{5.3}$$

Here π_k was only determined based on the different-distribution data. For reweighted-LDC the classifier was retrained using the updated weights w_i. This way the different-center samples that fit the modelled data better obtained a higher weight and contributed more than different-center samples that did not fit with the data which got a lower weight.

5.3 Experimental set-up

5.3.1 Image data

We used image data acquired within the multi-center PARISK study [205], which is a large prospective multi-center imaging study to improve risk stratification in patients with mild to moderate carotid artery stenosis. Inclusion criteria were a recent (<3 months) transient ischemic attack (TIA), amaurosis fugax or minor stroke, and a symptomatic carotid artery plaque of at least 2-3 mm with a stenosis <70% as determined on Doppler ultrasound or CT angiography. All patients underwent MRI, CT-angiography and ultrasound imaging of the carotid artery. For the present study we selected the first 20 patients from the Maastricht University Medical Center (center 1) and the first 22 patients from the Erasmus Medical Center (center 2), for whom a complete MRI session was available. MR imaging was performed on 3.0-Tesla whole-body scanners. Center 1 used an Achieva TX scanner (Philips Healthcare, Best, Netherlands) with an eight-channel phased-array coil (Shanghai Chenguang Medical Technologies Co., Shanghai, China), and center 2 used the Discovery MR 750 system (GE Healthcare, Milwaukee, MI, USA) with a four-channel phased-array coil with an angulated setup (Machnet B.V., Roden, the Netherlands).

The MRI protocol has been described previously [205], and is summarized in Table 5.1. The main differences between the two centers, apart from the differences in scanner model and coil, are the voxel sizes (both acquired and reconstructed), the use of a T1w IR-TFE (center 1) versus a SPGR scan (center 2) to visualize IPH, and a TOF FFE (center 1) versus a FSPGR (center 2) scan. This FSPGR sequence has been designed specifically to improve visualization of calcification in one image sequence to overcome the need to combine information from multiple sequences. For center 1 for all sequences 15 axial slices were acquired, while for center 2 the SPGR and FSPGR images were acquired in the coronal direction. Examples images from both centers are provided in Figure 5.1.

Table 5.1: MRI scan parameters.

Pulse sequence	T1w QIR TSE	T1w DIR FSE	TOF FFE	FSPGR	IR-TFE	SPGR	T2w TSE	T2w DIR FSE
	(Both pre- and postcontrast)							
Center	1	2	1	2	1	2	1	2
TR (ms)	800	1RR	20	3.3	9.1	9	4800	2RR
TE (ms)	10	5.2	5	2.1	5.5	1.3	49	50
FA (°)	-	-	20	5	15	30	-	-
Acquired voxel size (mm)	0.62×0.67	0.55×0.71	0.62×0.62	1.00×1.25	0.62×0.63	1.00×1.25	0.62×0.63	0.55×0.71
Reconstructed voxel size (mm)	0.30×0.30	0.55×0.63	0.30×0.30	0.63×0.63	0.30×0.24	0.63×0.63	0.30×0.30	0.55×0.63

* Abbreviations: TR = repetition time, TE = echo time, FA = flip angle, RR = R wave to R wave interval (1 heart beat), QIR = quadruple inversion recovery, TSE = turbo spin echo, DIR = double inversion recovery, FSE = fast spin echo, TOF = time of flight, FFE = fast field echo, FSPGR = fast spoiled gradient echo, IR = inversion recovery, TFE = turbo field echo, SPGR = spoiled gradient echo.

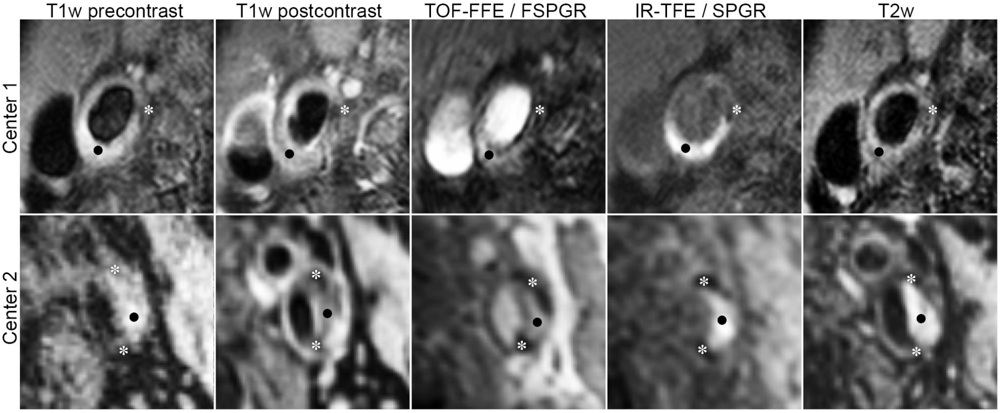


Figure 5.1: Examples of registered MR images from both centers. Center 1: A calcium spot (*) is indicated, which appears hypointense on all sequences. A region of IPH (black dot) is hyperintense on the IR-TFE sequence, TOF-FFE, and T1w precontrast, and shows no signal enhancement on the postcontrast T1w image. Center 2: a hyperintense region on the SPGR and T1w precontrast scan indicates IPH (black dot), and is hypointense on the T1w postcontrast scan. Two hypointense regions of calcification (*) are visible in especially the FSPGR and SPGR scans.

5.3.2 Manual reference

Manual contours of the symptomatic artery in each image were obtained for training and validation of the automatic methods. For center 1, the 20 scans were annotated by two observers with 3 years of experience with carotid MRI, using vessel wall analysis software (MRI-Plaque View, VPDiagnostics Inc, Seattle, USA, WA). First the images were semi-automatically aligned by registering the four other images to the T1w precontrast scan using the built-in registration tool [119]. This registration was manually adjusted for errors. Subsequently, lumen and outer vessel wall were semi-automatically [119] segmented and manually adjusted. Plaque components (CA, LRNC and IPH) were fully manually delineated, based on previously determined criteria [28, 30, 106] as agreed on by both observers on beforehand. IPH was defined as a hyperintense area in the IR-TFE scan, LRNC as a region that shows no contrast enhancement on the postcontrast T1w scan and is iso- or hyperintense on the precontrast T1w scan, and CA as hypointense on at least three image sequences. The remaining tissue within the vessel wall was considered as FT. All slices for which the five image sequences were available after co-registration, were annotated. One patient was excluded due to excessive patient movement.

As we used VesselMass (department of Radiology, Leiden University Medical Center, the Netherlands) for further analysis, editing and visualization, the annotated contours were converted into a format for use in VesselMass. Contour coordinates on the fixed T1w image could be extracted from the PlaqueView contours, and rigid registration with manual adjustments was repeated in VesselMass until the alignment of contours with images was determined to be correct by one of the observers. Due to significant inter-observer variability for this center, a consensus contour set was created by five experienced observers including the first two observers, with knowledge of the two initial segmentations, but without knowledge of any automatic segmentation results that were obtained using the first two contour sets. For this dataset segmentation and registration was approved by all observers.

The 22 scans from center 2 were manually annotated by one observer with 3 years of experience, using VesselMass. A subset of 10 scans was annotated by a second observer with 1 year of experience. First, the lumen was manually annotated on the T1w precontrast scan. Subsequently the remaining four image sequences were automatically registered to the T1w precontrast scan using a previously described algorithm for 3D rigid registration [220]. All contours (lumen, outer wall, CA, LRNC and IPH) were fully manually drawn. Similar to center 1, LRNC was defined as a hypointense area on the postcontrast T1w scan that is iso- or hyperintense on the precontrast T1w scan, and IPH as a hyperintense region on the SPGR scan. The criterium for CA segmentation was different from center 1: all hypointense regions in the FSPGR sequence were defined as calcium, without taking information from the other image sequences into account. In addition, areas with hypointensity in two or more of the other sequences without hypointensity in the FSPGR sequence were annotated as calcium if this was thought to be related to misregistration of the FSPGR volume. For both centers, by definition, all IPH lesions were drawn within a region of LRNC.

5.3.3 Data preprocessing

The images after registration as obtained and approved by the observers, were used for segmentation methodology development and evaluation. The scans acquired in center 2, especially the SPGR and FSPGR scans, showed a considerable intensity bias field due to coil inhomogeneity (Figure 5.1). This was corrected for in all five sequences by N4 inhomogeneity correction [206]. The images from center 1 did not show any coil inhomogeneity in the images, so N4 was not applied to the images from this center. Images from both centers were normalized in order to obtain similar intensity ranges between subjects. For image normalization a region of interest (ROI) was identified in each image. This ROI was defined as a 4×4 cm region around the lumen center in each image slice. The 5^{th} % of the intensity histogram in the 3D ROI per image was set to 0 and and the intensity of all voxels was linearly scaled such that the 95^{th} % of the histogram was set to 1000, for each scan individually. We assume this ROI was large enough to exclude any influence from plaque composition to the 5% and 95% histogram values.

5.3.4 Experiments

Three different training approaches for voxel classification were evaluated. For all approaches for center 1, the consensus contours were used for training and evaluation. For center 2 we used the contours of the observer who annotated all 22 datasets.

I) Same-center training: Methods were trained and evaluated on data from the same center, and thus acquired using the same hardware and imaging protocol. For both centers we performed leave-one-subject-out cross-validation.

II) Different-center training: Here a classifier developed on all vessels from center 1 was applied to segment the data from center 2, and vice versa, without the use of any labeled same-center samples during training.

III) Transfer learning: We simulated the situation in which a few slices from a larger set of same-center data are selected and manually segmented. This is practically feasible, and allows the use of transfer learning methods to tune the segmentation algorithm for use on the target data while most of the training data originates from the different-center dataset. In order to do this we selected a number of slices from both datasets that were considered suitable for training in such a setting. The selection criterium was presence of at least one of the three components CA, LRNC and IPH with a size of at least 10 voxels. This led to a selection of 118 out of the 285 slices for center 1 and 128 out of 359 slices for center 2. Experiments were performed with a random selection of 1, 3, 5 and 10 slices of target data, where those slices together contained at least 10 voxels of all components. Those slices were randomly selected from the other vessels from the target data in a leave-one-patient-out fashion. For each slice set selection, training was performed on these slices only, and weighted- and reweighted-LDC were performed with five different settings of weighting between different-center and same-center data. ΣW_{sc} was set to 0.1, 0.2, 1, 5 and 10, while ΣW_{dc} was always 1. To prevent effects of different components sizes due to slice selection, the prior probabilities for the classes (π_k) were set to the prior obtained from the fully annotated different-center dataset in all experiments. These experiments were repeated 100 times, to account for the variability between slice set selections. In each of the 100 iterations all vessels from the target data were segmented once.

5.3.5 Evaluation

Segmentation results were evaluated by 1) intraclass correlation coefficients for the volume of FT, CA, LRNC and IPH per vessel, 2) the error per component: the absolute difference between the amount of that component in the ground truth and the segmentation result per vessel, and 3) accuracy as % of correctly classified voxels. For same center training data confusion matrices were also used to enable a more localized analysis of segmentation accuracy. All experiments were performed both with and without feature normalization by adaptive histogram binning. In the transfer learning experiments with a limited number of annotated slices, the entire set of same-center data was used to determine the histogram bins and normalization of the individual slices was performed using these parameters. Results were compared with the inter-observer variability as determined between the two observers and the consensus reading for center 1, and between the two observers for center 2. For a more fair comparison of the segmentation results with the interobserver variability, the contours of the observers were evaluated within the consensus contours for the vessel wall for center 1, and for center 2 the contours of observer 2 were evaluated within the vessel wall contours of observer 1. Only voxels annotated within this wall were considered, and voxels not annotated within this wall were considered to be fibrous tissue. This is similar to how the automatic segmentation works, which also takes the reference vessel wall as an input.

Statistical comparisons were made between 1) same-center training, 2) different-center training, 3) different-center training with adaptive histogram binning, 4) training on a few same-center slices, 5) transfer learning and 6) transfer learning with adaptive histogram binning. The mean error of the four components for all 41 patients was taken. The analysis was done per center, and for the two centers combined, for the setting where 5 labeled same-center slices are available. For transfer learning we selected the method (weighted- or reweighted LDC and ΣW_{sc}) that overall performed best. For this method for each patient we took the median error over the 100 repeated experiments to use in the statistical analysis. Comparisons were made by Friedman analysis, followed by Tukey-Kramer testing for individual differences.

5.4 Results

Detailed results for volume error and ICC per component, and overall classification accuracy for each of the evaluated approaches are provided in Table 5.5 for center 1 and in Table 5.6 for center 2. Results for interobserver agreement are included as reference. It should be noted that for center 1 the accuracy of the observers is slightly biased since the consensus contours were based on these contours. In this section we will first present the results obtained in the traditional setting when a fully labeled dataset of same-center data is available, and secondly the results for different-center training and thereafter the results using transfer learning.

5.4.1 Same-center training

Figure 5.2 shows the correlations of tissue component volumes per vessel for the two centers, using same-center training without adaptive histogram binning. Good correlations were obtained for FT, CA and IPH for both centers. A good correlation for LRNC was obtained for the data from center 1, but a considerable underestimation with low correlation was obtained for center 2. Confusion matrices to assess voxelwise agreement are provided in Tables 5.2 and 5.3. These show a low sensitivity for LRNC (21% for center 1 and 12% for center 2) and a moderate sensitivity for CA (41% and 44%) which were both often misclassified as FT, a good sensitivity for IPH (66% and 75%) and high sensitivity for FT (97% both centers). Further results (volume errors and ICC values) for same-center training are summarized in the first part of Figures 5.3 and 5.4 (A-B). For center 1 (Figure 5.3), using adaptive histogram binning slightly increased performance on average, but showed a slight decrease for LRNC and IPH. For center 2 for same-center training adaptive histogram binning had a positive effect on LRNC segmentation, and did not influence the other components.

When volume errors and ICC values are compared with the interobserver agreement (Tables 5.5-5.6), similar accuracies with respect to the consensus contours were found for the observers and for the automated same-center methods for center 1. For center 2, the ICC for LRNC was slightly lower than the agreement between observers, and for CA slightly higher, but errors for FT, LRNC and CA were larger.

5.4.2 Different-center training

The second part of Figures 5.3 and 5.4 (C-D) provides results for different-center training. For center 1 training on only different-center data resulted in an extreme overclassification

Table 5.2: Confusion matrix for center 1 Same-center training without adaptive histogram binning

		Classification result			
		FT	LRNC	CA	IPH
Ground truth	FT	87%	0.7%	2.1%	0.4%
	LRNC	2.0%	0.7%	0.2%	0.3%
	CA	2.4%	0.1%	1.7%	0.0%
	IPH	0.5%	0.4%	0.0%	1.8%

Table 5.3: Confusion matrix for center 2 Same-center training without adaptive histogram binning

		Classification result			
		FT	LRNC	CA	IPH
Ground truth	FT	80.6%	1.0%	1.4%	0.3%
	LRNC	2.6%	0.4%	0.1%	0.2%
	CA	4.1%	0.0%	3.4%	0.3%
	IPH	0.9%	0.4%	0.0%	4.1%

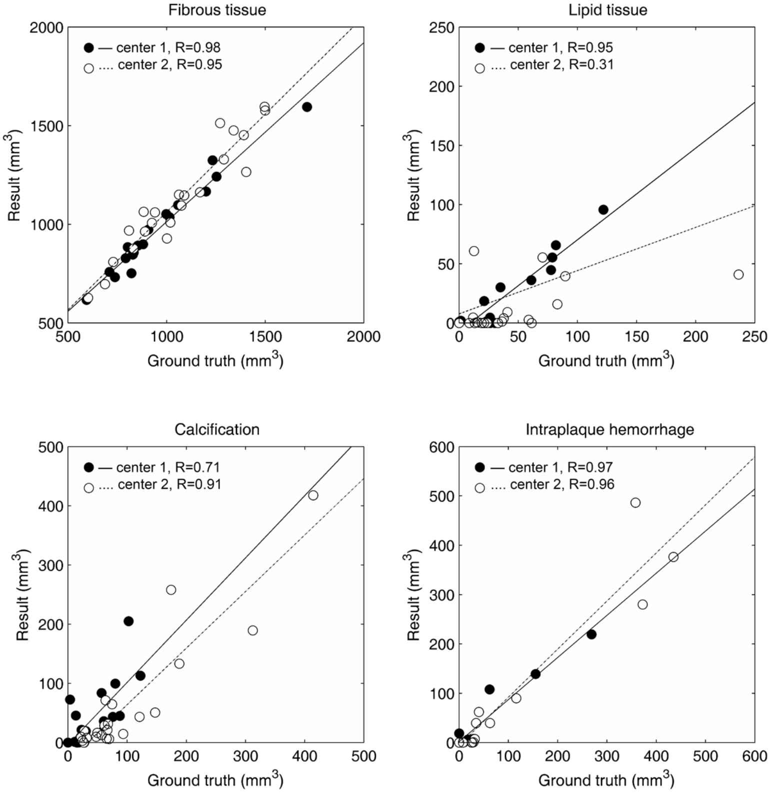


Figure 5.2: Correlations for same-center training results without adaptive histogram binning. Volumes per vessel are shown for the two centers. Pearson correlation R-values were calculated for the two centers separately.

of FT as CA, resulting in large errors for these components. Accuracy for LRNC also decreased, but IPH segmentation remained equally accurate as for same-center training. The errors were much lower when adaptive histogram binning was used. However, slightly larger errors remained present for FT, CA and IPH compared with same-center training. For center 2 segmentation accuracy was much less influenced by training on different-center data. Classification of CA was slightly negatively affected, but the accuracy for the other components did not change much. For center 1 different-center training without adaptive histogram binning yielded significantly larger errors than same-center training, but different-center training with adaptive histogram binning did not. For center 2 this was the other way around: different center training with adaptive histogram binning performed significantly worse than same-center training, but direct different-center training did not. In the combined analysis different-center training both with and without adaptive histogram binning had significantly larger errors than same-center training.

5.4.3 Transfer learning

Results for the transfer-learning experiments are summarized in Figure 5.5 (center 1) and Figure 5.6 (center 2). Figure 5.5 shows that without adaptive histogram binning better

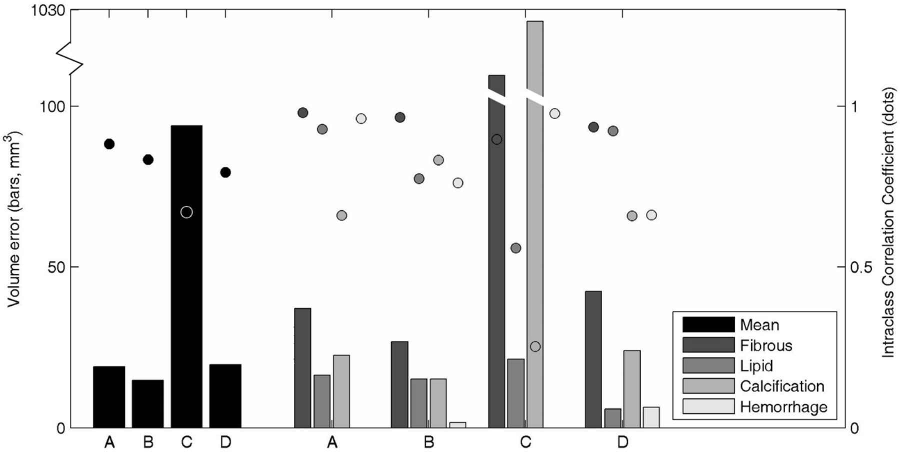


Figure 5.3: Errors and intraclass correlation obtained for center 1. A) Training on same-center data, B) training on same-center data including adaptive histogram binning, C) training on different-center data and D) training on different-center data including adaptive histogram binning. Volume errors (bars) are given as the median over all 19 vessels, the black bars represent the mean of the four medians. For hemorrhage no bars are shown in A and C, as in 11 of the 19 vessels no hemorrhage was present both in the ground truth and segmentation result. For reference, total vessel wall volume for this center was 1057±301 ml.

results were obtained using transfer learning than for different-center training (see Figure 5.3). While for weighted-LDC the ratio between ΣW_{dc} and ΣW_{sc} does not seem to have a large influence, for reweighted-LDC it is apparent that a smaller weight on the same-center data deteriorated the results. This was the case both with and without adaptive

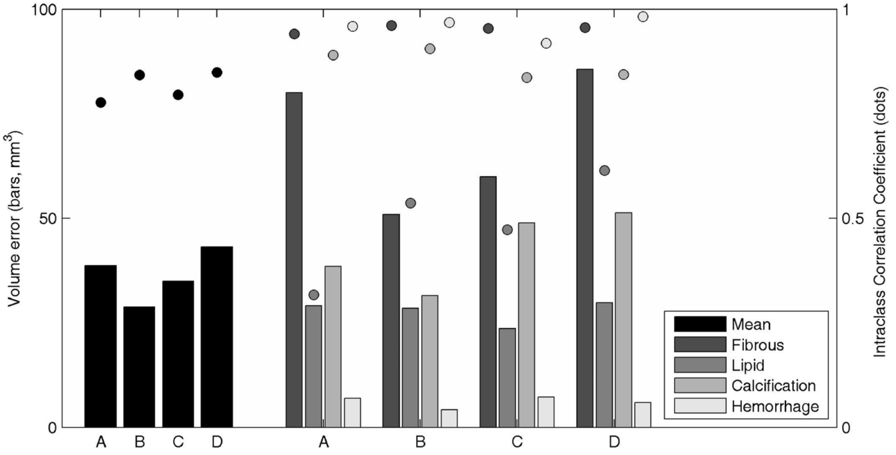


Figure 5.4: Errors and intraclass correlation obtained for center 2. A) Training on same-center data, B) training on same-center data including adaptive histogram binning, C) training on different-center data and D) training on different-center data including adaptive histogram binning. Volume error (gray bars) is given as the median over all 22 vessels, the black bars represent the mean of the four medians. For reference, total vessel wall volume for this center was 1340±389 ml.

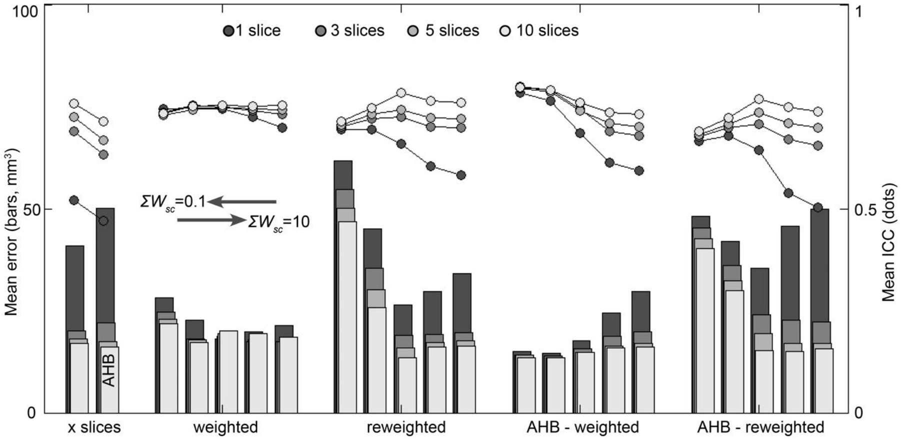


Figure 5.5: Mean error and intraclass correlation (ICC) for center 1 when only a limited amount of same-center training data is available. For each set of 5 transfer learning experiments ΣW_{sc} from left to right is 0.1, 0.2, 1, 5, 10. AHB = adaptive histogram binning. The bars and dots for 'x slices' are the result for training only on the labeled same-center slices.

histogram binning. Adaptive histogram binning in combination with weighted-LDC yielded the best results. In Table 5.5 results for separate components are given for the example with 5 annotated slices from the same-center data, and ΣW_{sc}=1. Additionally, the results for weighted-LDC with ΣW_{sc}=0.1 and adaptive histogram binning are provided, as this yielded the lowest mean error and highest mean ICC. It can be seen that for all transfer-learning methods the overall accuracy was higher than for different-center training.

While for center 2 training on different-center data already yielded reasonable results, improvement could still be obtained by transfer learning. As for center 1, reweighting was more sensitive to the ratio between ΣW_{dc} and ΣW_{sc}. However, here a smaller weight on the different-center data yielded more accurate results than when ΣW_{sc} increased. Best results were obtained for combining adaptive histogram binning and sample reweighting. In Table 5.6 results for separate components are given for the example with 5 annotated slices from the target data, and ΣW_{sc}=0.1.

We performed statistical analysis for weighted-LDC and ΣW_{sc}=0.1. For both centers, and the centers combined, both transfer learning with and without adaptive histogram binning of the features the errors were not significantly different from same-center training. The difference with training on 5 same-center slices was not significantly different either. Additionally, from the figures it can be seen that for almost all approaches for both centers the use of more annotated target slices yields lower mean volume errors and higher correlation values. In Table 5.4 the average number of target samples and the % of the combined source and target data is given for both centers. The difference is due to a difference in voxel size between center 1 (0.30×0.30×2 mm) and center 2 (0.55×0.63×2 mm). This percentage, however, did not influence the importance of the same-center samples during training as these were determined by the chosen ΣW_{sc}, and the class priors which were determined by the different-center data only.

Table 5.4: Percentage of labeled same-center samples in the total training set for transfer learning experiments.

	center 1 n samples (% of total)	center 2 n samples (% of total)
1 slice	721±200 (2%)	253±90 (0.2%)
3 slices	1736±359 (4%)	580±131 (0.5%)
5 slices	2790± 461 (6%)	935±169 (0.8%)
10 slices	5492±637 (11%)	1831±236 (1.6%)

5.4.4 Visualization of results

Segmentation results for three slices from both centers are shown in Figure 5.7. For both centers the results for transfer learning were obtained using 5 randomly selected same-center slices. Both used adaptive histogram binning on the features and ΣW_{sc}=0.1, but for center 1 the results using weighted-LDC are shown, and for center 2 using reweighted-LDC. Of the 100 repeated experiments with random selection of 5 target slices, for each specimen we used the selection for which the total error over the four components was closest to the median total error of the 100 experiments for the examples shown.

The segmentations show that same-center training usually yields the smoothest results. Different-center training and transfer learning yield slighly more noisy results, but similar overall segmentations as same-center training. The transfer-learning segmentations have a slightly better detection of CA and IPH than different-center training in some of the examples.

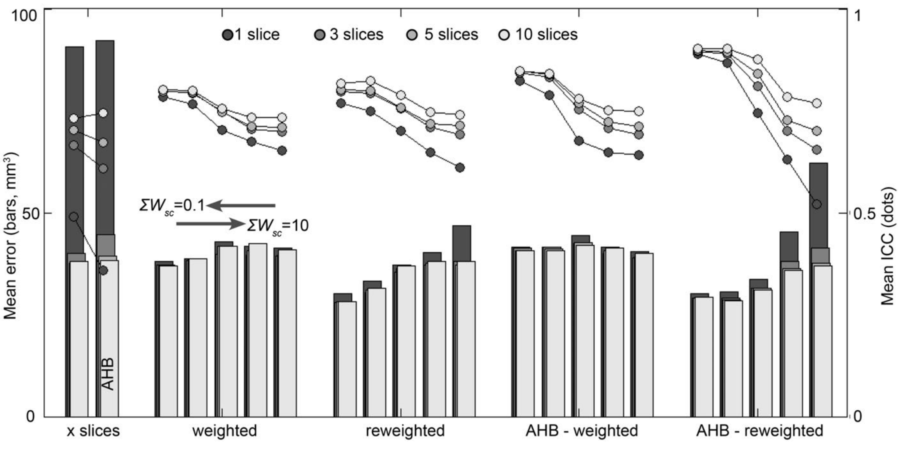


Figure 5.6: Mean error and intraclass correlation (ICC) for center 2 when only a limited amount of same-center training data is available. For each set of 5 transfer learning experiments ΣW_{sc} from left to right is 0.1, 0.2, 1, 5, 10. AHB = adaptive histogram binning. The bars and dots for 'x slices' are the result for training only on the labeled same-center slices.

Table 5.5: Segmentation results for center 1.*

	Median error mm^3 (IQR)				ICC (95% Confidence interval)				Accuracy (%)
	FT	LRNC	CA	IPH	FT	LRNC	CA	IPH	(Median (IQR))
Obs1 - Obs*	58 (25-72)	34 (12-75)	24 (11-37)	1 (0-23)	0.96 (0.89-0.99)	0.34 (-0.17-0.71)	0.42 (-0.1-0.75)	0.19 (-0.31-0.62)	89 (85-94)
Obs1*	41 (21-61)	20 (13-35)	17 (7-28)	0 (0-16)	0.97 (0.92-0.99)	0.75 (0.42-0.91)	0.87 (0.67-0.95)	0.26 (-0.25-0.66)	90 (89-94)
Obs2	14 (7-26)	10 (5-32)	7 (1-20)	0 (0-14)	0.99 (0.96-0.99)	0.71 (0.39-0.88)	0.81 (0.56-0.92)	0.92 (0.82-0.97)	94 (92-97)
Automated									
Same center	37 (9-64)	16 (1-24)	23 (12-33)	0 (0-15)	0.98 (0.65-0.99)	0.93 (0.83-0.97)	0.66 (0.31-0.85)	0.96 (0.90-0.99)	91 (90-95)
Same center AHB	27 (13-50)	15 (3-29)	15 (7-25)	2 (0-24)	0.96 (0.91-0.99)	0.77 (0.50-0.91)	0.83 (0.62-0.93)	0.76 (0.48-0.90)	93 (89-95)
Other center	169 (97-204)	21 (4-33)	185 (143-218)	0 (0-6)	0.90 (0.75-0.96)	0.56 (0.15-0.80)	0.25 (-0.22-0.63)	0.98 (0.94-0.99)	74 (70-79)
Other center AHB	42 (18-78)	6 (1-13)	24 (10-35)	6 (1-24)	0.93 (0.84-0.97)	0.92 (0.81-0.97)	0.66 (0.30-0.85)	0.66 (0.31-0.85)	87 (83-92)
*Transfer Learning **									
5 target slices	35 (15-65)	17 (5-28)	20 (9-39)	1 (0-14)	0.96 (0.90-0.98)	0.59 (0.20-0.82)	0.56 (0.16-0.81)	0.88 (0.72-0.95)	90 (86-94)
5 target slices, AHB	30 (12-67)	16 (4-28)	18 (9-36)	6 (1-27)	0.94 (0.86-0.98)	0.59 (0.20-0.82)	0.57 (0.17-0.81)	0.54 (0.13-0.80)	90 (85-94)
Weighted, ΣW_{sc}=1	37 (19-64)	21 (3-30)	20 (10-33)	0 (0-9)	0.98 (0.95-0.99)	0.55 (0.15-0.80)	0.63 (0.26-0.84)	0.97 (0.93-0.99)	92 (88-95)
Weighted, ΣW_{sc}=1, AHB	24 (10-58)	15 (2-26)	15 (7-27)	5 (0-26)	0.95 (0.87-0.98)	0.62 (0.24-0.83)	0.69 (0.35-0.87)	0.72 (0.40-0.88)	91 (86-94)
Weighted, ΣW_{sc}=0.1, AHB	29 (10-57)	10 (2-19)	11 (4-26)	6 (0-28)	0.94 (0.85-0.98)	0.85 (0.66-0.94)	0.69 (0.35-0.87)	0.69 (0.35-0.87)	90 (85-93)
Reweighted, ΣW_{sc}=1	29 (12-65)	13 (4-25)	20 (9-44)	1 (0-16)	0.95 (0.86-0.98)	0.65 (0.28-0.85)	0.52 (0.09-0.78)	0.90 (0.75-0.96)	89 (85-94)
Reweighted, ΣW_{sc}=1, AHB	35 (12-78)	12 (3-24)	18 (8-38)	13 (3-36)	0.92 (0.80-0.97)	0.73 (0.43-0.89)	0.61 (0.23-0.83)	0.63 (0.25-0.84)	88 (83-93)

* All values shown are with respect to the consensus reading. Interobserver variability is calculated within the annotated consensus vessel wall contours. For observer 1 contours of 3 vessels were lost for conversion to VesselMass format and were not used. For transfer learning all results are given as the median over the 100 experiments where 5 labeled same-center slices were used. FT = Fibrous tissue, LRNC = Lipid-rich necrotic core, CA = Calcification, IPH = Intraplaque hemorrhage, IQR = Interquartile range, ICC = Intraclass correlation coefficient, AHB = Adaptive histogram binning.

Table 5.6: Segmentation results for center 2.*

	Median error mm^3 (IQR)				ICC (95% Confidence interval)				Accuracy (%) (Median (IQR))
	FT	LRNC	CA	IPH	FT	LRNC	CA	IPH	
Observer 2 *	22 (17-37)	8 (7-27)	18 (8-30)	5 (0-18)	0.97 (0.86-0.99)	0.76 (0.29-0.93)	0.75 (0.28-0.93)	0.98 (0.92-0.99)	88 (85-91)
Automated									
Same center	80 (42-126)	29 (14-53)	38 (22-67)	7 (0-28)	0.95 (0.89-0.98)	0.30 (-0.13-0.64)	0.91 (0.80-0.96)	0.96 (0.90-0.98)	90 (84-95)
Same center AHB	51 (28-93)	28 (9-40)	32 (12-44)	4 (0-24)	0.98 (0.94-0.99)	0.66 (0.33-0.84)	0.90 (0.78-0.96)	0.98 (0.95-0.99)	90 (85-95)
Other center	60 (37-92)	24 (10-50)	49 (28-71)	7 (0-31)	0.97 (0.94-0.99)	0.56 (0.19-0.79)	0.85 (0.68-0.94)	0.92 (0.83-0.97)	88 (80-94)
Other center AHB	86 (53-130)	30 (11-50)	51 (27-69)	6 (0-21)	0.97 (0.94-0.99)	0.76 (0.50-0.89)	0.85 (0.67-0.93)	0.99 (0.97-0.99)	89 (85-95)
Transfer Learning									
5 target slices	71 (33-114)	24 (10-53)	48 (25-85)	8 (0-31)	0.93 (0.84-0.97)	0.34 (-0.09-0.66)	0.66 (0.34-0.84)	0.88 (0.73-0.95)	87 (80-93)
5 target slices, AHB	76 (36-137)	23 (9-46)	47 (25-89)	12 (1-34)	0.89 (0.76-0.95)	0.37 (-0.05-0.68)	0.52 (0.14-0.77)	0.77 (0.52-0.90)	87 (79-93)
Weighted, ΣW_{sc}=0.1	67 (41-100)	22 (9-51)	50 (25-69)	8 (0-30)	0.97 (0.93-0.99)	0.54 (0.16-0.78)	0.84 (0.66-0.93)	0.93 (0.83-0.97)	88 (81-95)
Weighted, ΣW_{sc}=0.1, AHB	82 (47-131)	25 (11-53)	49 (27-70)	6 (0-24)	0.98 (0.94-0.99)	0.73 (0.45-0.88)	0.87 (0.71-0.94)	0.99 (0.97-0.99)	90 (86-95)
Reweighted, ΣW_{sc}=0.1	53 (27-125)	15 (6-33)	36 (17-71)	9 (0-30)	0.92 (0.83-0.97)	0.75 (0.48-0.89)	0.76 (0.50-0.89)	0.93 (0.85-0.97)	87 (79-92)
Reweighted, ΣW_{sc}=0.1, AHB	54 (28-89)	15 (6-33)	37 (19-65)	10 (1-28)	0.96 (0.91-0.98)	0.91 (0.79-0.96)	0.87 (0.71-0.94)	0.97 (0.93-0.99)	89 (84-94)

* Interobserver variability is only on 10 vessels. Contours from observer 2 are considered within the annotated vessel wall by observer 1. For transfer learning all results are given as the median over the 100 experiments where 5 labeled same-center slices were used. FT = Fibrous tissue, LRNC = Lipid-rich necrotic core, CA = Calcification, IPH = Intraplaque hemorrhage, IQR = Interquartile range, ICC = Intraclass correlation coefficient, AHB = Adaptive histogram binning.

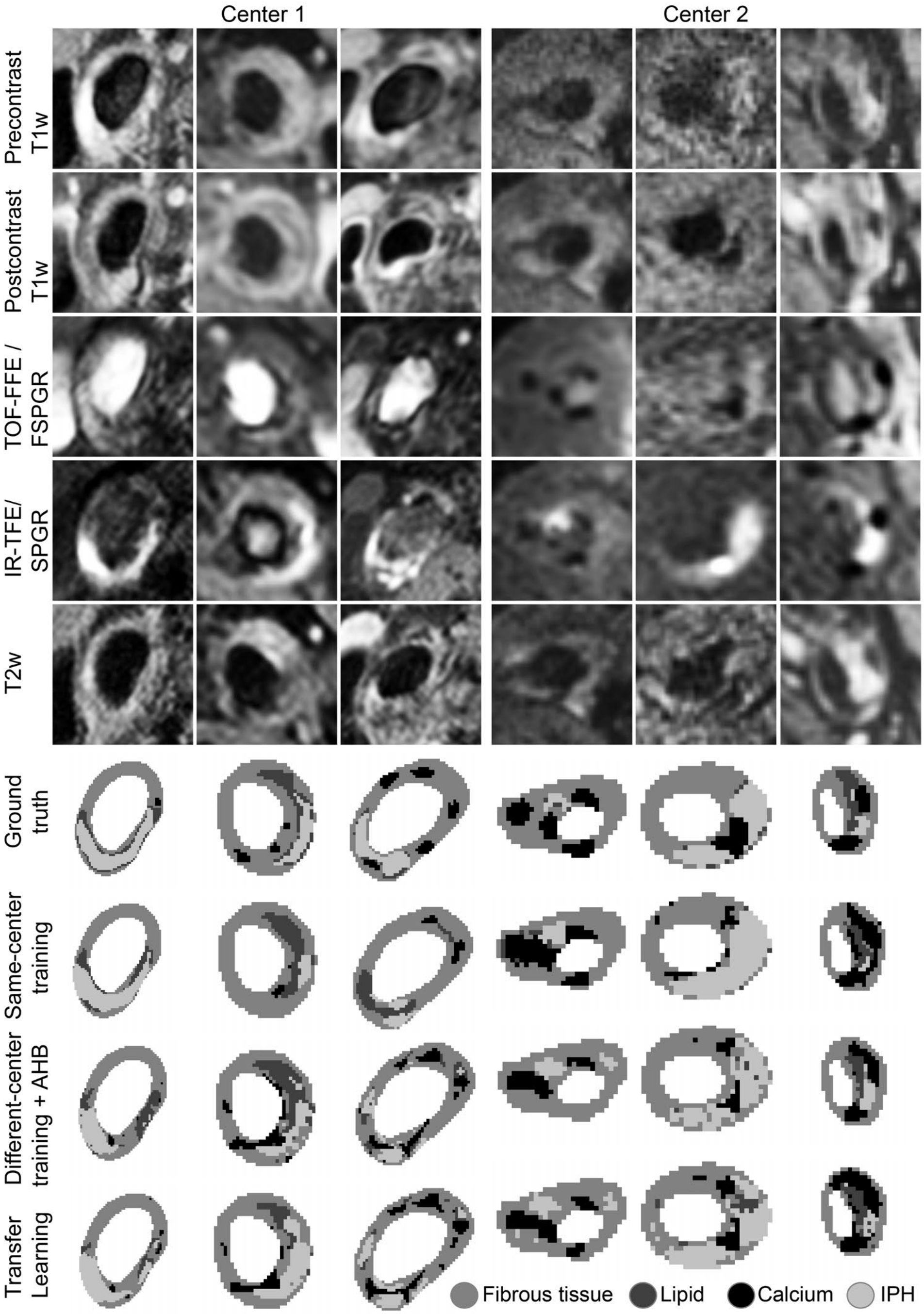


Figure 5.7: Segmentation results for three slices from center 1 and three slices from center 2. The results for transfer learning were obtained using 5 selected target slices and both adaptive histogram binning and weighted-LDC with ΣW_{sc}=0.1 (center 1), or AHB and reweighted-LDC with ΣW_{sc}=0.1 (center 2).

5.5 Discussion and conclusion

In this work we performed plaque-component segmentation in a multi-center MRI study. For both centers the obtained segmentations showed a similar agreement with manual annotations as the interobserver agreement. Moreover, we showed that both extensive feature normalization using adaptive histogram binning, and transfer-learning algorithms can improve segmentation accuracy when only few annotations from a certain target dataset are available and a much larger annotated dataset with different feature distributions is used.

Applying a classifier optimized using training data from one center to image data from the other center yielded large errors for center 1, but reasonable results for center 2. The largest errors were obtained for CA, which can be explained by the large differences in image acquisition and manual analysis for this component. In center 2 hypointense regions in the FSPGR scan were annotated as CA, while for center 1 the corresponding TOF-FFE has a bright lumen but overall dark vessel wall, resulting in large CA overestimation when using the classifier developed for center 2. Appearance of IPH and LRNC was more similar between centers and raised fewer problems. For center 1 the accuracy for CA was significantly improved using adaptive histogram binning, but this decreased accuracy for IPH. This also occurred for same-center training, so this is probably related to difficulties in modeling the IPH class after adaptive histogram binning for this center. LDC is optimal if all classes follow a Gaussian distribution with equal covariance, which cannot be assumed after the performed histogram stretching. For center 2 adaptive histogram binning on the different-center data yielded a larger overall error compared to direct different-center training.

Transfer learning could improve over different-center training with or without adaptive histogram binning, with only the need to obtain labels for a small number of slices (1-10). For both centers a combination of adaptive histogram binning and transfer learning performed best. However, the best transfer-learning algorithm and settings differed between the centers. For center 1, where training on the different-center data yielded large errors, with reweighted-LDC an equal contribution of ΣW_{dc} and ΣW_{sc} performed best, while with adaptive histogram binning and weighted-LDC a larger weight on the different-center data performed slightly better. For segmentation in center 2 training on the different-center data without adaptive histogram binning already performed well, and giving large weights to the different-center data performed best for all transfer-learning algorithms. The higher ΣW_{sc} is set, the more importance is given to a small amount of samples which may hamper accurate modeling. Here reweighted-LDC does perform slightly better than weighted-LDC. Although the best performing algorithm differed between the two centra, overall, weighted-LDC with or without adaptive histogram binning and ΣW_{sc}=0.1 performed best and did not differ significantly from same-center training.

In future studies it can be useful to perform an initial evaluation to get an indication of what method and settings would work. It is possible to first train a classifier solely on the different-center (or eg. different-scanner) data, and visually inspect results on some examples from the same-distribution data. If this performs reasonably well, weighting the source data higher than the target data is safe, and reweighting may improve performance as for center 2 in this study. If, on the other hand, different-center training yields large errors, a more equal weighting between source and target data should be considered and reweighting should be performed with more caution. There is a risk that reweighting increases the weight of source samples that fit well with the source-data distribution but that are not representative for the target samples. In all our experiments, reweighting with a large weight on the same-center samples had low performance, but still similar or better than using only the same-center slices.

There are certain requirements to be able to use both adaptive histogram binning and transfer learning. One requirement to successfully use adaptive histogram binning is that a representative set of data needs to be available for both centers, such that the distribution of disease stage and therefore prior class probabilities is similar. For both adaptive histogram binning and the used transfer-learning algorithms it is important that there is a direct link between each feature in one dataset with one of the features in the other dataset. This can be obtained by linking features from two sequences with comparable contrast properties. In the multi-center study that we used, the MRI protocol was designed to be comparable between the two centers; only the image data was different due to use of a different scanner and institutional preferences. If no obvious link is present, either only the sequences that are comparable in both datasets can be used, or measures such as the Kullback-Leibler divergence of the histograms can be used to determine which sequences are most similar in appearance. It is also possible to combine two or more sequences from one dataset to be similar to a sequences in the other dataset by summing them with different weights such that the divergence from this target sequences in the other dataset is minimal. In addition, it is assumed that the ordering of classes is the same for all features for the different centers. In our dataset this did not fully hold for the TOF-FFE/FSPGR sequence, where IPH had the highest intensity in the FSPGR scan, but the second-highest after FT in the TOF-FFE scan. However, if sufficient weight is given to the same-center data, such features will contribute less to the LDC as LDC aims to optimize the discrimination between classes.

The suitability of adaptive histogram binning and the used transfer-learning algorithms may depend on which classifier is used. Adaptive histogram binning distorts Gaussianity of features which is assumed by LDC, but using adaptive histogram binning did not negatively affect our results for same-center training, except for IPH segmentation in center 1. Almost all classifiers can deal with sample weighting. However, our reweighting procedure was more tuned to LDC. For other classifiers, such as SVM, different criteria for reweighting have been used [216] which may be more suitable for that specific classifier.

In our data there was a slight difference in prior probabilities for the tissue components between the two centers, which could have influenced the adaptive histogram binning in different-center training. More CA (8% vs 4%) and IPH (5% vs 3%) was manually segmented for center 2. This could be due to a difference in patient population and vessel wall composition in the included patients, but likely also results from differences in the imaging protocol. Although this has not been studied in detail, it seems the FSPGR sequence to image CA in center 2 leads to an overestimation of CA, like the blooming effect that is seen in CT. Use of this dedicated sequence to image CA did on the other hand yield larger ICC values for CA than the traditional MR protocol used in center 1.

The performance of transfer learning depends on the suitability of the selected slices. These need to be representative of all classes in the target data. We used a minimum of 10 voxels for each class as selection criterium for the total set of selected slices, where each slice has at least one component with at least 10 voxels besides FT. In practice this is not expected to raise problems, as slices with large areas of CA, LRNC or IPH can easily be recognized by human observers. In our experiments the number of selected slices was more important for center 1 where training on different-center data yielded larger errors. For center 2 the overall error did not differ much between using 3 or 10 slices, but the average ICC was slightly better when more same-center slices were used. The number of same-center training samples in the smallest class was significantly related with the average error for both centers and all transfer-learning experiments combined (R=-0.05, $p<10^{-32}$). Although the error by training on 5 same-center slices only was not significantly different from same-center training or transfer learning, it should be noted that the priors here were also determined on the different-center data as the presence of CA, LRNC and IPH would be overestimated based on the selected slices.

The proposed method for feature normalization is similar to Nyul et al. [138], except that they perform histogram matching per image instead of the features of a large group of subjects combined. We chose this approach, since the tissue distribution in the vessel wall differs more between patients the tissue distribution of the brain. Compared to previous literature on plaque-component segmentation, our methods have similar [119] or slightly better [82, 219, Chapter 3] accuracy than previously published results on same-center training and evaluation. Our results imply that such methods can more easily be implemented in multi-center studies, although standardization of image protocols remains advantageous for obtaining a link between image sequences. In previous studies LRNC and IPH have sometimes been considered together as one ‘vulnerable’ component. If we would combine our segmentations for these components, correlations would be slightly higer (center 1) or close to (center 2) the results for IPH only.

In conclusion, good plaque-component segmentations with similar accuracies as inter-observer agreement were obtained for the data from both centers. We showed that when no labeled same-distribution data is available extensive feature normalization by means of adaptive histogram binning improves results, and secondly that transfer-learning classifiers can improve results when a few labeled same-center examples are available. These approaches yield results with similar accuracy to same-center training. Both feature normalization and transfer learning can facilitate segmentation across scanners, and facilitate wide implementation of automated image analysis methods in large-scale multi-center studies and in clinical practice.

Acknowledgements

We thank Zaid Kassar for his manual annotations, and Floris Schreuder for his assistance in creating the consensus segmentation.

Chapter 6

Three-dimensional carotid ultrasound plaque texture predicts vascular events

This chapter is based on:
A. van Engelen, T. Wannarong, G. Parraga, W.J. Niessen, A. Fenster, J.D. Spence and M. de Bruijne, Three-dimensional carotid ultrasound plaque texture predicts vascular events, *Submitted*

* * *

Abstract
Background and purpose: Carotid ultrasound atherosclerosis measurements including those of the arterial wall and plaque provide a way to monitor patients at risk of vascular events. Our objective was to examine carotid ultrasound plaque texture measurements and the change in carotid plaque texture over 1 year in patients at risk of events, and to compare these with measurements of plaque volume and other risk factors as predictors of vascular events in patients at risk of events.
Methods: We evaluated 298 patients with carotid atherosclerosis using three-dimensional ultrasound at baseline and after one year and measured carotid plaque volume and 376 measures of plaque texture. Patients were followed up to 5 years for myocardial infarction, TIA and stroke. Sparse Cox Regression was used to select the most predictive plaque texture measurements in independent training sets using a tenfold cross-validation, repeated five times, to ensure unbiased results.
Results: Receiver-operator-curves and Kaplan-Meier analysis showed that changes in texture and TPV combined provided the best predictor of vascular events. In multivariate Cox regression changes in plaque texture (median HR=1.4, $p<0.001$) and TPV (median HR=1.5 per 100 mm^3, $p<0.001$) were both significant predictors, whereas the Framingham risk score was not.
Conclusions: Changes in both plaque texture and volume are strongly predictive of vascular events. In high-risk patients 3D ultrasound plaque measurements should be considered for vascular event risk prediction.

6.1 Introduction

There is an urgent need for rapid, reliable and cost-effective methods to monitor patients who are at high risk for major adverse vascular events. In other words, new ways to monitor patients at risk are important to target treatment to the right patients for preventing vascular events [171]. Ultrasound (US) is a relatively inexpensive, and widely available imaging method for quantitative measurements of the carotid artery wall including intima-media thickness (IMT), vessel wall volume, plaque burden and both echogenic and echolucent plaque. For example, ultrasound measurement of IMT is widely used to measure arterial wall thickness abnormalities. Although in some populations it predicts vascular risk, IMT and IMT progression do not strongly predict vascular events [47, 122]. On the other hand, carotid plaque burden measures such as total plaque area (TPA) or total plaque volume (TPV) and their changes over time may provide stronger predictors of events [229]. It has been shown that these can be reliably and relatively easily measured [4, 111].

In addition to the presence and volume of carotid plaque, the presence of a large lipid core or intraplaque hemorrhage is believed to destabilize plaque, while calcifications have a stabilizing effect [104, 160, 167, 168, 188]. It is in this context that carotid US measurements of plaque texture analysis have been developed. A relationship between the amount of lipid present in a plaque and US texture has been determined [51, 236]. Most studies have focused on the degree of echogenicity of plaques using the fact that lipid and hemorrhagic areas are more echolucent while calcified and fibrous areas are echorich [110]. US echogenicity has been shown to differentiate between symptomatic and asymptomatic subjects [64] and has been used to predict events [69, 94, 116, 135, 149]. An approach that incorporates higher-order texture parameters may provide more insight into the underlying tissue properties and has been used in several studies as well [2, 35, 109]. In previous studies, more complex, higher-order texture measures were shown to differentiate accurately between symptomatic and asymptomatic subjects [2], and performed better than a set of plaque shape parameters [35]. In addition, they provided a higher predictive value for events than a combination of a history of events and plaque parameters such as plaque area and gray scale median [109].

In addition to single time-point measurements, progression of TPV was a strong predictor of events [229], and changes in plaque texture were more sensitive to statin-induced effects than changes in TPV [16]. Based on all these findings, we hypothesize that plaque volume change over time is related to changes in plaque composition and plaque vulnerability and perhaps subsequent vascular events. Therefore our goal is to develop a novel methodology for three-dimensional US (3DUS) plaque texture measurements and determine: 1) if plaque texture and changes in plaque texture are predictive of vascular events and 2) if the combination of changes in plaque texture and volume improve prediction of events in high-risk patients.

6.2 Materials and Methods

Study population, image acquisition and plaque volume annotation have previously been described in Wannarong et al. [229].

6.2.1 Study Population

Patients with a history of risk factors such as hypertension or hyperlipidemia, or with a history of vascular events, that were being followed up in the Stroke Prevention Clinic or the Premature Atherosclerosis Clinic at the University Hospital, London, Canada, were enrolled in the study. The inclusion criteria included a baseline plaque area between 40 and 600 mm^2, measured by 2DUS as previously described [182]. Participants with a stenosis ≥70% as measured by Doppler US were excluded and all subjects provided written informed consent to a protocol approved by the Western University Human Ethics Research Board.

6.2.2 Follow-up for Outcomes

Participants were followed up to five years. They were annually questioned about vascular events during the previous year, and events reported were confirmed by review of the hospital electronic record. If participants had died, death and cause of death were confirmed by the primary care physician. Myocardial infarction, TIA, stroke, revascularization (stenting, bypass, endarterectomy of any artery) and death were reported.

6.2.3 Ultrasound Acquisition

3D US scans (most common voxel size $0.21 \times 0.21 \times 0.21$ mm, mean $0.21 \times 0.21 \times 0.33$ mm) of both carotid arteries were acquired of all patients at baseline and after one year; the duration of followup was (median [range]) 364 [226-897] days. The ultrasound transducer (L12-5, 50 mm, Philips, Bothel, WA) was moved along the neck of the patient for 4.0 cm in 8 seconds, imaging ~30 slices/s. Video frames were digitized using 3DEchotec equipment (General Electric Medical Systems, Halbergmoos, Germany) coupled to the ATL HDI 5000 scanner. The acquired 2D images were reconstructed immediately into a 3D US image (3DQuantify, Robarts Imaging Laboratories, London, Canada).

6.2.4 Total Plaque Volume

For plaque delineation the 3D US image was displayed using 3D viewing software (3DQuantify, Robarts Research Institute, Western University, London, Canada). TPV was semi-automatically measured [27, 111] in all scans by one observer (T.W.) who was blinded to the time point (baseline or follow-up). For each plaque, a plane in the longitudinal view which best visualized the plaque was selected for defining the end points of the plaque. At 25%, 50% and 75% of its length perpendicular to the long axis the boundary of the plaque was annotated. These boundaries, together with the two end points, were used to create a volume by creating a polyhedron that connects the boundary points. The TPV of all plaques, of both arteries, per patient were added to obtain one value per patient per time point.

6.2.5 Texture Analysis

A set of 376 texture measures from 9 different texture techniques was calculated. Most of the texture measures have previously been used in studies on 2D [2, 35, 109] and 3D [16] carotid US. Our parameter settings were based on these studies, but texture measures are calculated for all three imaging planes, or for the complete plaque volumes. An overview is provided in Table 6.1. The texture measures were implemented in Matlab (Matlab R2011B, The MathWorks Inc., Natick, MA). More details can be found in the appendix (page 84).

Table 6.1: A list of texture measures that were used in the analysis. A more detailed description can be found in the appendix (page 84)

Method	Number of features
Gray level distribution [16]	34
Gray-level co-occurrence matrix [75]	78
Gray-level run length matrix [62]	60
Gray-level difference matrix [234]	12
Neighbourhood gray tone difference matrix [10]	10
Laws texture [113]	105
Local Binary Pattern [140]	27
Gaussian Filter Bank [179]	24
Structure Tensor [233, Chapter 4]	20

Before calculating texture measures, the images were normalized. For each image the 10^{th} and 90^{th} percentile of the imaged volume were fixed and intensities were linearly scaled with these values.

Texture was calculated for the annotated plaque volumes in both arteries. For each measure the average of the left and right artery was calculated by weighting the two values by the plaque volumes, to obtain one value per patient for each measure, at both baseline and follow-up subsequently. For seven cases with no plaque in the 3D ultrasound at baseline, baseline texture was taken as the mean of all other subjects at baseline. Texture measures were normalized by setting the mean of all subjects to 0 with a standard deviation of 1 for each measure.

Sparse Cox regression was used to combine the 376 texture measures into one *texture-based risk indicator*, using the glmnet toolbox [58, 172] for R [150]. With sparse regression, a penalty term promotes the reduction of the number of parameters in the model, leaving only the strongest predictors. In our experiments we fixed the number of remaining parameters in the model to five, due to the relatively low number of events.

To ensure unbiased parameter selection, experiments were performed by 10-fold cross-validation, so subjects were divided over 10 equal-sized groups with the event incidence equally divided over those 10 folds. Within cross-validation experiments the model for combining texture measures was built on 9 of the 10 folds, and used to calculate the hazard ratio (HR) for the subjects in the 10^{th} fold. This was repeated 10 times with each fold left out once, to obtain an HR for each subject. This HR was used as the texture-based risk indicator. This was carried out both for the 376 baseline texture parameters, and for 376 texture change parameters, calculated by subtracting baseline texture from follow-up texture.

To combine two or more parameters (texture/TPV baseline/changes) a Cox model was built using only those parameters. This was done using the same 10-fold cross-validation: both the model to combine the texture measures into one risk parameter, and the model to combine parameters was developed on 9 folds and evaluated on the 10^{th}. All cross-validation experiments were repeated five times, to evaluate stability of the procedure.

6.2.6 Statistical analysis

Due to non-normality of the studied parameters continuous variables are given as median and interquartile range (IQR), and Kruskal-Wallis testing was used to test for differences between groups. For categorical variables a χ^2 test was used to compare groups.

The predictive value of the studied parameters was evaluated by receiver-operating characteristic (ROC) analysis, Kaplan-Meier analysis, and Cox regression where besides texture change and TPV change a number of clinical parameters were included. For ROC analysis, the area under the curve (AUC) was determined for baseline texture, texture change, baseline TPV and TPV change and combinations of those four measures. To test for statistical differences the HRs of the 5 repetitions were averaged for all parameters, except TPV and TPV change. ROC curves were compared using DeLong tests for correlated curves. Further analyses focused on change in texture and TPV, and the combination of those two. Kaplan-Meier curves were made after dividing the participants in tertiles: low-risk, medium-risk and high-risk subjects, based on ranking subjects by increasing TPV or texture change (HR). The difference between the tertiles was tested for statistical significance by log rank tests. In addition, the hazard ratio of the high-risk group with respect to the other two groups was determined by Cox regression. In this regression analysis a binary indication of high-risk versus low-or-medium-risk was used rather than continuous values.

To evaluate the potential effects of the ultrasound-derived texture measure in combination with covariates that were previously shown to be predictive, Cox regression was performed on all patients with predictors change in TPV, change in plaque texture and the extended Framingham risk score [40]. The extended Framingham risk score predicts general cardiovascular disease risk, based on age, sex, HDL and total cholesterol, systolic blood pressure (treated or untreated), smoking and diabetes. Due to skewness of the texture measure, the logarithm of this measure was used. The full model was compared to the model without texture change, and the model without TPV change, using a log-likelihood ratio test.

In all experiments participants who experienced MI, TIA, stroke or death due to vascular events during followup were considered as positive for vascular events. Statistical analysis was performed using both R (Sparse Cox regression, ROC and Kaplan Meier analysis) and SPSS software (full Cox regression models and likelihood ratio tests).

6.3 Results

6.3.1 Patient characteristics

In total 349 patients were enrolled in the study. After excluding patients with at least one unsuccessful US, or neither measurable plaque in 3DUS at baseline nor at follow-up, 298 patients were eligible for US analysis. During follow-up (median [range] 3.12 [0.77-4.66]) years) 27 of these subjects experienced a vascular event, of which nine had a stroke, eleven a Transient Ischemic Attack and seven a myocardial infarction. Two subjects died due to the adverse event. Baseline patient characteristics, for patients with and without an event during follow-up are provided in Table 6.2.

6.3.2 Texture measures

In Figure 6.1, two of the strongest texture measures (when considering change) are shown for a single patient. These two measures were most often selected by Cox regression: 49 and 48 out of 50 times. The two plaques in this vessel are different in appearance, which is reflected by the two texture images. Table 6.3 provides the six most often selected texture measures as selected by sparse Cox regression, alongside with texture change and TPV

Table 6.2: Baseline characteristics

	No event (n=271)∗	With event (n=27)∗	p-value†
Age	70 (64-77)	75 (64-80)	0.30
Male	58%	52%	0.54
Systolic blood pressure (mmHg)	134 (122-149)	131 (114-142)	0.11
Diastolic blood pressure (mmHg)	74 (66-82)	69 (62-78)	0.17
TPV (mm^3)	273 (191-437)	253 (119-422)	0.21
Stenosis (%)	40 (40-50)	40 (40-40)	0.26
Body Mass Index	27.9 (25.5-31.5)	29.0 (25.3-31.8)	0.96
Smoking (never, quit, still smoking)	36% - 56% - 8%	33% - 48% - 19%	0.74, 0.45, 0.06
Smoking pack-years	5 (0-24)	15 (0-28)	0.34
Total cholesterol, mmol/L	3.9 (3.4-4.7)	4.1 (3.5-4.9)	0.35
HDL cholesterol, mmol/L	1.3 (1.1-1.7)	1.3 (1.1-1.8)	0.85
LDL cholesterol, mmol/L	1.9 (1.5-2.5)	2.1 (1.5-2.6)	0.30
Triglycerides, mmol/L	1.2 (0.9-1.7)	1.1 (0.9-1.8)	0.77
Diabetes Mellitus	21%	19%	0.79
Previous stroke	22%	33%	0.19
Previous TIA	42%	56%	0.18
Previous MI	14%	26%	0.11
Previous endarterectomy	7%	4%	0.48
Previous carotid angioplasty	1%	4%	0.26
Previous peripheral artery angioplasty	1%	4%	0.26
Previous coronary angioplasty	7%	11%	0.44
Coronary Artery Bypass Graft (CABG)	10%	15%	0.47
Taking antihypertensive drug	89%	93%	0.52

∗ Continuous variables are given as median (interquartile range); categorical variables are given as percentages.

† p-values using Kruskal-Wallis testing for continuous variables and using χ^2 testing for categorical variables.

change, and their values in patients who do or do not experience a vascular event for reference. In practice only these strongest texture parameters need to be calculated, upon which the risk parameter can be determined and compared with reference values.

6.3.3 ROC analysis

In Figure 6.2A four ROC curves are shown for baseline TPV (AUC=0.57) and TPV change (0.72), and for baseline texture (0.68±0.01) and texture change (0.74±0.02) averaged over five repeated cross-validation experiments. Figure 6.2B shows ROC curves when two or more features are combined: AUC is 0.71±0.02 for the combination of baseline and change in TPV, 0.75±0.03 for baseline and change in texture, 0.66±0.02 for baseline texture and TPV, 0.78±0.02 for changes in texture and TPV, 0.78±0.01 for baseline texture and change in TPV, 0.79±0.02 for baseline texture and changes in texture and TPV, and 0.78±0.02 for all 4 parameters: baseline and change in texture and baseline and change in TPV.

After averaging over 5 repetitions, the AUCs of baseline and change in texture were significantly different ($p<0.05$), as were texture change and TPV change ($p<0.01$). The combination of texture and TPV changes was, however, not significantly different from texture change ($p=0.24$). Still, for further results we focus on TPV change, texture change, and their combination.

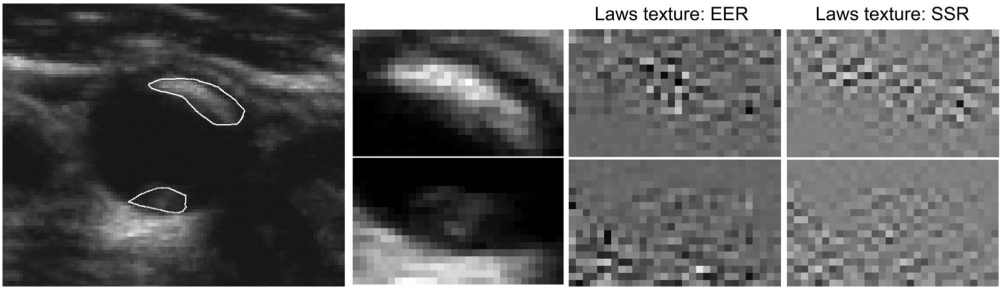


Figure 6.1: Texture for two plaques in the same vessel with a different appearance. In a total of 50 runs of sparse Cox regression (5 times 10-fold cross-validation) on changes in texture, Laws EER was selected in the model 49 times, and Laws SSR 48 times.

6.3.4 Kaplan-Meier

Figures 6.3-6.5 show Kaplan-Meier curves for TPV change, texture change and their combination. The difference between the high-risk and low-risk tertiles was significant for all experiments ($p=0.01$ for TPV change, $p\leq0.003$ for texture change and $p\leq0.001$ for the combination), the difference between the medium-risk and high-risk tertile in 3 out of 5 repetitions for texture change ($p\leq0.012$) and in all repetitions for the combination ($p\leq0.012$), and the difference between the low-risk and medium-risk in 1 out of 5 repetitions for both texture change ($p<0.05$) and the combination ($p<0.05$).

Comparing the high-risk tertile to the combined low-risk and medium-risk tertiles with Cox regression, showed a HR of 2.3 ($p=0.03$) for TPV change, 4.3 (2.7-6.4) (median, range) with $p<0.001$-0.01 for texture change, and 6.2 (4.2-7.9) with $p<0.001$ for the combination.

6.3.5 Cox regression

Table 6.4 shows the results for Cox regression with TPV change, texture change and the Framingham risk score. Both TPV change and texture change were significant predictors

Table 6.3: Overview of the texture measures that were determined to be most predictive, for the groups with and without events during follow-up.

	No event (n=271) (median(range))	With event (n=27) (median(range))	p-value∗
TPV change (mm^3)	21 (-53 – 77)	80 (29 – 170)	0.0002
Texture change: mean of Laws EER	-0.13 (-0.61 – 0.41)	0.50 (0.08 – 1.26)	0.0005
Texture change: NGTDM Coarseness (bins of 1)	-0.06 (-0.63 – 0.53)	0.49 (0.02 – 1.20)	0.0013
Texture change: mean of Laws SSR	0.11 (-0.31 – 0.42)	-0.35 (-0.85 – -0.05)	0.0004
Texture change: mean of Laws SSS	0.00 (-0.38 – 0.47)	-0.53 (-0.96 – 0.06)	0.0002
Texture change: NGTDM Contrast (bins of 1)	0.02 (-0.27 – 0.29)	-0.32 (-1.33 – 0.15)	0.0061
Texture change: GLCM in axial plane, standard deviation of correlation (bins of 10)	0.13 (-0.53 – 0.58)	-0.56 (-1.08 – 0.04)	0.0014
Texture change: risk parameter †	0.97 (0.82 – 1.14)	1.24 (1.05 – 1.46)	0.00005

∗ Kruskal-Wallis test

† HR after combining 5 measures of texture change. The given result is the median over the five experiments with different cross-validation folds.

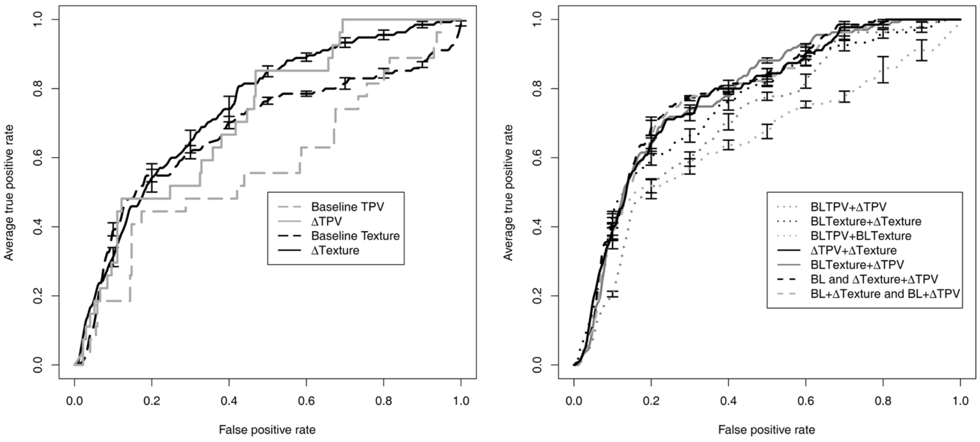


Figure 6.2: ROC curves for the prediction of vascular events. Black bars represent the standard error over 5 repetitions. A) Area under the curve (AUC) is 0.0.57 for baseline TPV, 0.68±0.01 for baseline texture, 0.72 for TPV change and 0.74±0.02 for texture change. B) AUC is 0.71±0.02 for the combination of baseline TPV and TPV change, 0.75±0.03 for baseline texture and texture change, 0.66±0.02 for baseline texture and TPV, 0.78±0.02 for texture and TPV change, 0.78±0.01 for baseline texture and TPV change, 0.79±0.02 for baseline texture and texture and TPV change, and 0.78±0.02 for all 4 parameters: baseline texture and TPV and texture and TPV change. Legend: BL=Baseline, Δ=Change.

of stroke, TIA and MI, while in this dataset the Framingham risk score was not. The log-likelihood ratio tests indicated that the full model was significantly better than the model excluding texture change, and the model excluding TPV change (all $p<0.01$). Subanalysis in Table 6.5 shows that when only stroke and TIA are considered as events, TPV change and texture change remain the only significant parameters in the model.

6.4 Discussion

This study compared the predictive value for vascular events of texture and TPV derived from 3D carotid ultrasound. The change in 3DUS texture characteristics over time was a

Table 6.4: Results for Cox regression. For texture change and for the full model the hazard ratios and p-values are given with median and range over 5 repetitions of calculating the texture-based risk indicator.

	Individual model		Full model	
Parameter	p-value (median(range))	Hazard ratio (median(range))	p-value (median(range))	Hazard ratio (median(range))
Texture change (per 0.1 change in log(HR))	<0.001	1.4 (1.3-1.5)	≤0.002	1.4 (1.3-1.5)
TPV change (per 100 mm^3)	<0.001	1.5	≤0.001	All 1.5
Framingham risk score	0.31	1.1	0.17 (0.14-0.19)	All 1.1

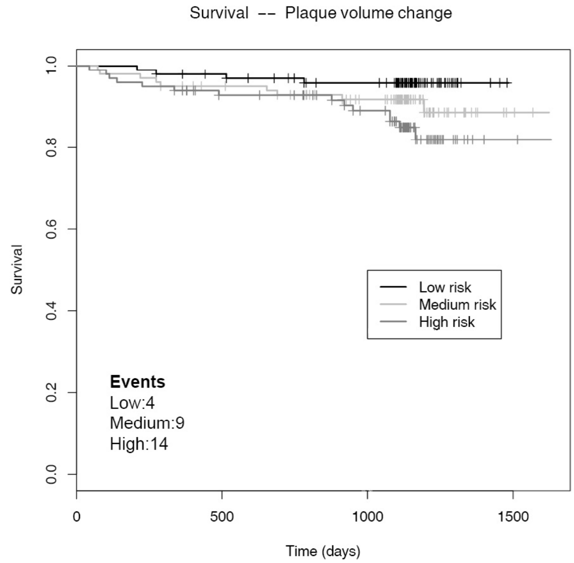


Figure 6.3: Kaplan-Meier curves for event-free survival for three tertiles of TPV change. The number of events per tertile, as median (range), is given for the 3 different models. P-value for high-risk vs. low risk is 0.010, for high-risk vs. medium-risk 0.227 and for low-risk vs. medium-risk 0.165.

stronger independent predictor than both baseline texture and changes in TPV over one year. Both texture and TPV change remained significant predictors after adjustment for the Framingham risk score. The full model was significantly better than the model including only TPV or texture change and the Framingham risk score.

The predictive value of TPV change [229], and baseline texture [109] have both been investigated previously, but to our knowledge this is the first study to consider changes in texture and to combine texture change and TPV change for risk stratification of patients over time. Texture change was more predictive of events than baseline texture, which suggests that plaques that are changing faster impose a higher risk on patients than plaques that are stable in texture and by analogy, stable in composition. This notion is also supported by the observation that in our analysis, for all texture measures that were selected, the median change is closer to zero for patients not experiencing an event than for those who do (Table 6.3).

In a previous study where statin-induced changes were better reflected by texture change than by TPV change [16], Laws texture measures performed best, which is similar to our

Table 6.5: Results for Cox regression in a subanalysis in which only stroke and TIA were considered as positive end points for events. The hazard ratio and p-value for all parameters are given with median and range over 5 repetitions of calculating the texture-based risk indicator.

Parameter	p-value (median(range))	Hazard ratio (median(range))
Texture change (per 0.1 change in log(HR))	≤0.003	1.4 (1.3-1.5)
TPV change (per 100 mm^3)	≤0.007	1.5 (1.4-1.5)
Framingham risk score	0.87 (0.83-0.94)	All 1.0

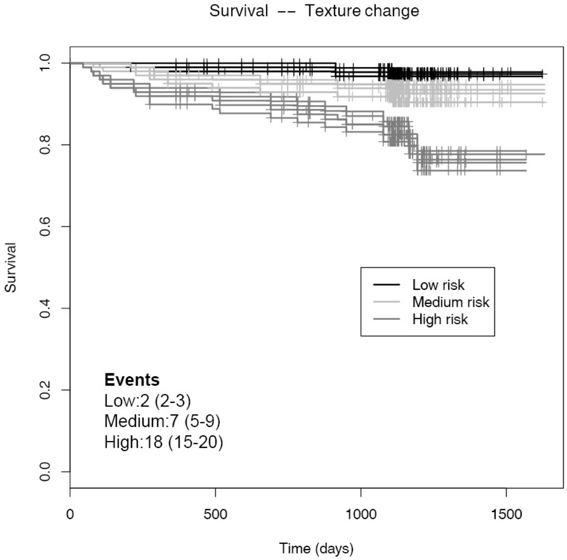


Figure 6.4: Kaplan-Meier curves for event-free survival for three tertiles of texture change. The 5 curves show the 5 repetitions of the experiments with texture. The number of events per tertile, as median (range), is given for the 3 different models. P-value for high-risk vs. low risk is ≤0.003 for all repetitions, for high-risk vs. medium-risk 0.001-0.151 and for low-risk vs. medium-risk 0.033-0.332.

findings. Moreover, NGTDM coarseness was important in our study, and was among the best measures in previous work that showed the discrimination between symptomatic and asymptomatic patients [35].

Our findings can be used both for patient monitoring and evaluation of therapies. The present study shows that including plaque texture in the monitoring of patients contributes to improved risk assessment using ultrasound. A yearly follow-up including ultrasound in an atherosclerosis clinic is practically feasible [183]. More regular follow-up of high-risk patients would be possible to enable adjustment of therapy in a more timely fashion. Adjusting patient treatment based on changes in plaque area instead of traditional risk factors was already shown to significantly reduce the number of vascular events. [181]. Including changes in TPV and texture could further improve effectiveness of patient management. In addition, cost-effective measurements are needed to evaluate newly developed therapies, where imaging parameters will be a good secondary endpoint as an alternative to the number of events for which larger studies and longer follow-up are required.

When myocardial infarction was left out as an endpoint, no relevant changes in the hazard ratio of both texture change and TPV change were observed in Cox regression. Although the number of myocardial infarctions was too small to study the effect for MI only, this suggests that carotid artery US may be used topredict risk for stroke/TIA and MI. This supports the notion that vascular disease is systemic, and that carotid atherosclerosis may accurately estimate and reflect disease in other vessels [90, 185].

A limitation of this study is the relatively small number of events. However, the finding that TPV change and texture change are both significant predictors of events even in this relatively small dataset supports the use of such US-derived measures compared to traditional risk factors. It should be noted that these results hold only for subjects with

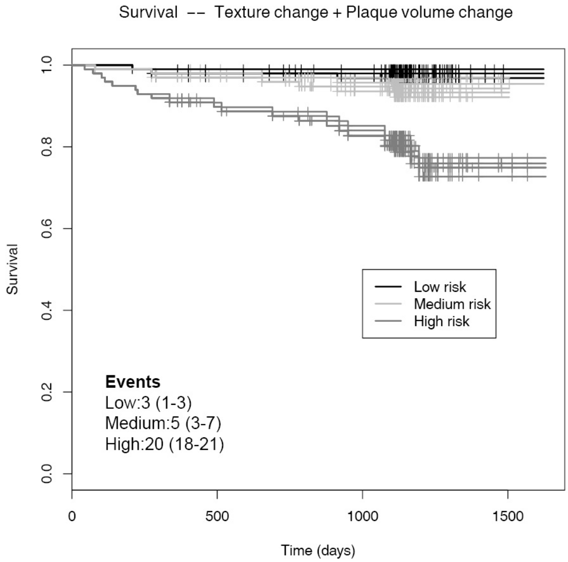


Figure 6.5: Kaplan-Meier curves for event-free survival for three tertiles of TPV change and texture change combined. The 5 curves show the 5 repetitions of the experiments with texture. The number of events per tertile, as median (range), is given for the 3 different models. P-value for high-risk vs. low risk is ≤0.001 for all repetitions, for high-risk vs. medium-risk ≤0.012 and for low-risk vs. medium-risk 0.031-0.999.

identified carotid atherosclerosis, which is a clinically relevant population. Additionally, the sensitivity of the proposed method to the used imaging equipment and settings has to be evaluated. Finally, it is important to acknowledge that in order to incorporate plaque texture into clinical practice and large clinical studies, automated analysis will be required. In the present study, plaque volumes were segmented semi-automatically, but automatic methods are also available [33] which will help facilitate translation to clinical practice.

6.5 Conclusion

Change in US plaque texture and volume are predictors of vascular events in patients in whom traditional risk factors are not. These measures can be used in clinical practice as a cost-effective way to monitor high-risk subjects, or as an evaluation measure in the development of new therapies.

Acknowledgements
Dr. J. David Spence was the principal investigator, and Dr. Daniel Hackam and Dr. Andrew House were co-investigators, in the study in which the images and data on risk factors and outcomes were obtained (reference [229]).

Appendix: Texture descriptors

Here we provide a more detailed description of the texture measures that have been included in the study. For each group of measures we summarize the settings that were used and provide a complete list of these measures. Before calculating the measures the images were normalized by setting the 10^{th} percentile of the intensity in the imaged volume to 10 and the 90^{th} percentile to 150. Texture was calculated for all 3D plaque regions in both carotid arteries, and the texture of the left and right artery were averaged by taking the weighted average with respect to plaque size, to obtain one value per patient for each measure.

Gray level distribution (GLD, 34 measures)

GLD measures were calculated from all voxels within the plaque regions. Our selection of measures was based on Awad et al. [16]. The selected measures were mean, standard deviation, median, minimum, maximum, entropy, mode, energy, the first 7 standardized moments around the mean, a normalized histogram of 20 bins ranging from 0-500, and the histogram bin with the highest count.

Gray-level co-occurrence matrix (GLCM, 78)

GLCM measures as introduced by Haralick et al. [75] measure the joint probability of pairwise combinations of image gray levels. In this paper co-occurences with a distance of 1 pixel were calculated in horizontal, vertical, and diagonal orientations (θ = 0, 45, 90, 135) for intensity bins with width 10. For the axial, coronal and saggital plane one co-occurence matrix was made that included all slices, and the mean and standard deviation of the 4 orientations were calculated per plane for 13 measures [75, 178]: autocorrelation, contrast, correlation, cluster prominence, dissimilarity, energy, entropy, homogeneity, maximum probability, sum average, sum entropy and two information measures of correlation

Gray level run length (GLRL, 66)

GLRL considers ‘runs’, which are defined as a set of consecutive pixels with the same gray level value [62] and of which the length is the number of pixels in that run. For the present paper, runs with θ = 0, 45, 90 and 135 were calculated in the axial, coronal and saggital plane. For each plane the following measures were calculated and averaged over the four orientations: Short run emphasis, long run emphasis, gray level non-uniformity, run length non-uniformity, run percentage, low gray level run emphasis, high gray level run emphasis, short run low gray level emphasis, short run high gray level emphasis, long run low gray level emphasis and long run high gray level emphasis. All measures were calculated for both intensity bins of width 5 and 20 and implemented using the toolbox by Wei [232].

Gray level difference matrix (GLDM, 12)

GLDM [234] calculates the absolute difference between neighbouring pixels. We used a distance of 1 pixel, and calculated the mean over 4 orientations (θ = 0, 45, 90, 135), for 3 planes (axial, coronal, saggital). For each plane the mean, contrast, angular second moment and entropy were calculated.

Neighbourhood gray tone difference matrix (NGTDM, 10)

NGTDM was developed to calculate five measures related to human perception: coarseness, contrast, busyness, complexity and strength [10]. We calculated the mean difference with the surrounding voxels in a 3D 26-neighbourhood, for intensity bins of 1 and 10, and calculated the five measures for both these settings.

Laws texture (105)

Laws developed texture energy measures by convolution of the image with 1D kernels [113]. We combined five 1D kernels to study 3D texture: [1 4 6 4 1] (Level, L), [-1 -2 0 2 1] (Edge, E), [-1 0 2 0 -1] (Spot, S), [1 -4 6 -4 1] (Ripple, R) and [-1 2 0 -2 1] (Wave, W), which created 125 3D kernels: LLL, LLE, etc. After convolution, we calculated the mean, absolute mean and standard deviation over the region of interest, and averaged measures of rotated kernels, such as LLE, LEL and ELL, or LES, LSE, ELS, ESL, SEL and SLE. This gave 35 kernels with three measures each.

Local Binary Pattern (LBP, 27)

The original 2D implementation of LBP measures the homogeneity of texture by determining the number of transitions from intensities higher than each central pixel to intensities lower than that central pixel [139, 140]. We applied a 3D adaptation where we determined the number of ‘regions’ higher or lower in intensity around each voxel in a 26- and a 98-neighbourhood. Measures were the mean and standard deviation of the number of regions in both the 26- and the 98-voxel neighbourhood, a normalized histogram of 1 up to 6 and >6 areas for the 26-neighbourhood, and 1 up to 15 and >15 areas for the 98-neighbourhood.

Gaussian filter bank (24)

We applied 3D Gaussian filters to calculate blurred intensity, gradient magnitude, Laplacian and curvature at 3 scales: 0.16 mm, 0.32 mm and 0.64 mm. As measures we used the mean and standard deviation of these measures over the regions of interest [179].

Structure tensor (20)

The structure tensor [233] measures coherence in an image. From the 3D structure tensor we calculated the fractional anisotropy, the 3 eigenvalues and the determinant, and calculated the mean and standard deviation over the regions of interest [Chapter 4]. We applied two different scales: an inner scale of 0.16 mm with an outer scale of 0.48 mm, and an inner scale of 0.32 mm with an outer scale of 0.80 mm.

Chapter 7

Summary and Discussion

In this thesis a variety of image analysis techniques for characterization of the atherosclerotic vessel wall in multimodal images have been developed and evaluated. Here we provide a summary of the main results, followed by a discussion and view on future perspectives.

7.1 Summary of contributions

7.1.1 Plaque characterization with histological reference

The first part of this thesis (Chapters 2-3) focused on the development of automated segmentation algorithms for plaque components in the carotid arteries, by using registered histology and μCT as a ground truth for training and evaluation. Compared to previous studies on plaque component segmentation that use histology as a reference, we used registration with a 3D histology stack instead of manual matching of selected 2D slices. This allows for a more accurate alignment since the rotation angle between MR imaging and histology sectioning can be accounted for. A previously proposed registration framework [68] for creating this histology stack, and 3D registration with *ex vivo* MRI and μCT, and *in vivo* CTA was used for the studies in this thesis.

Chapter 2 made use of this dataset and focused on the development of voxel classification methods for segmenting *ex vivo* MRI. The images features distances to the wall, Gaussian smoothing, gradient magnitude and Laplacian at multiple scales, improved segmentation accuracy compared to using only normalized MR intensities. Good results for quantification of calcification, lipid-rich necrotic tissue and fibrous tissue were obtained with respect to histology and μCT. The classifier using all features performed significantly better than previously used subsets of features. In addition, we obtained slightly higher accuracies using 3D registration than using 2D slice-matching and registration, suggesting that a more accurate registration by taking rotation into account may be important for accurate classifier training. The difference was not statistically significant, which may have been caused by the limited number of specimens and the relatively small imaging angle between *ex vivo* MRI and histology compared to *in vivo* MRI.

In Chapter 3 we performed plaque component segmentation in a combination of *in vivo* MRI and CTA. The same methodology and 3D-registered dataset as in Chapter 2 were used. We added registration of *in vivo* MRI to the framework presented in Groen et al. [68], and made a few small improvements to the framework that improved the overlap of the *in vivo* data with histology. However, still no perfect voxelwise correspondence, which is needed for training an accurate classifier, was obtained. To compensate for this, we developed and evaluated three ways of sample weighting for classifier training. When all samples were used without taking registration errors into account, good correlations between estimated relative plaque components volumes and the ground truth were found, but the size of the lipid-rich necrotic core was largely underestimated. Using the LDC classifier, two approaches decreased this bias: 1) when the ground truth segmentations were blurred to obtain soft labels indicating uncertainty at tissue boundaries, combined with sample weighting based on the Dice overlap of the vessel wall in histology and the *in vivo* data per slice, and 2) with Gaussian outlier rejection. However, only outlier rejection decreased the absolute error of the segmented component volumes. With non-linear SVM weighting samples individually based on local registration error measured as local contour distance yielded best results. The overall best method was LDC with Gaussian outlier rejection, which yielded results with good accuracy for volume quantification of all studied components.

In addition we showed that a combination of CTA, which can accurately segment calcification, and MRI, which accurately discriminates between lipid and fibrous tissue, yields good results for all three tissue types. Since CTA is commonly used to image symptomatic patients when they enter the hospital, and MRI can better image plaque components related to increased vulnerability, it can be advantageous in clinical practice to image these high-risk patients with both CTA and MRI.

7.1.2 Histology segmentation

Chapters 2 and 3 relied on manual histology segmentations for ground truth generation. This is a very time-consuming process, which is why in Chapter 4 we focused on automatically segmenting histology sections. We performed fully automatic segmentation by training on independent data, and semi-automatic segmentation for which we trained a specimen-specific classifier on 1 or 2 selected slices for each specimen. Fully automatic segmentation yielded good results for fibrous and lipid segmentation, but low sensitivity for necrotic tissue. Semi-automatic segmentation resulted in increased sensitivity for necrotic tissue, whose appearance varied the most between different specimens. However, the overall accuracy was not significantly different between automatic and semi-automatic segmentation.

These (semi-) automatic segmentations were used to train algorithms to segment *ex vivo* MRI and *in vivo* MRI and CTA data as presented in Chapters 2 and 3. For lipid-rich necrotic core volumes large errors were found in both *ex vivo* and *in vivo* data. This can be attributed to undersegmentation of necrotic tissue in the automatically segmented histology sections. However, the semi-automatically segmented histology sections yielded equally good MRI segmentations as manual histology segmentations.

7.1.3 Multi-center MRI segmentation

Whereas in Chapters 2-4 we used one dataset that was created with the same protocol, in Chapter 5 we studied the accuracy of plaque component segmentation in a multi-center study with carotid MRI scans acquired in two different medical centers. We used manual MRI annotations instead of histology sections as the ground truth reference. We used the same classifier and Gaussian filter bank MRI features as in Chapter 3, only the acquired MR sequences were not exactly the same. When training two separate classifiers for the two centers by training on labelled same-center data, for both centers we obtained segmentations of fibrous tissue, lipid tissue, calcification, and intraplaque hemorrhage with similar agreement with a manual segmentation as the inter-observer agreement.

Directly applying a classifier trained on data from one center to data from the other center performed well in one direction, but yielded large errors in the reverse direction. We evaluated two approaches to improve this transferability of plaque-component classifiers developed on data from one center or scanner, to data from a different source. Firstly, an extensive feature normalization method that performs histogram stretching by adaptive histogram binning, largely improved transferability in the case where directly applying the different-center classifier with only image intensity normalization gave large errors. Secondly, we applied transfer learning algorithms, and assumed that a small set of manually labelled slices from the same-center data was available. These algorithms perform weighted classification in which samples from same-center data receive a larger weight than different-data samples. This yielded an overall error that was not significantly different from same-center training.

7.1.4 Ultrasound texture analysis

In Chapter 6 we focused on the analysis, and predictive value for vascular events, of 3D carotid ultrasound. Ultrasound is widely available and relatively cheap, so it is more feasible for widespread clinical use. In this study we did not segment plaque components individually, but instead characterized plaque vulnerability by studying the average texture of all plaques present. A large set of texture features was extracted for patients with known carotid atherosclerosis, at two time points with an interval of one year. Using cross-validation the most predictive features for vascular events (stroke, TIA, myocardial infarction) were selected by sparse Cox regression. We showed that changes in texture are more predictive than baseline texture, and that the combination of change in texture with change in plaque volume improves prediction of risk for vascular events. Although this study was performed in a relatively small cohort of 298 patients, ultrasound-derived features were shown to be a better risk-indicator than the Framingham risk score for general vascular disease.

7.2 Discussion

In this discussion we first discuss the contributions of the work described in this thesis, in relation to each other and to the literature. We will then look a bit further and put our contributions in perspective of related topics that were not directly part of this thesis, and discuss future perspectives regarding both methodological developments and the translation to clinical practice.

7.2.1 Accuracy of plaque component segmentation

Plaque components that were studied in this thesis, i.e. calcification, lipid/necrotic core, and intraplaque hemorrage, as determined from MRI or CT, are known to be related to plaque vulnerability [70, 104]. We achieved calcification classification with very high accuracy in *ex vivo* MRI in Chapter 2, but only with very low accuracy in *in vivo* MRI in Chapter 3, and reasonable accuracy in Chapter 5. A possible explanation for the difference in performance is that calcification spots are relatively small. High-resolution *ex vivo* images can accurately image those spots, while *in vivo* imaging suffers more from partial volume effects. In addition, in Chapters 2 and 3 the ground truth was obtained from registered histology data. Misregistrations have a larger effect on these smaller structures, which can severely complicate classifier training. In Chapter 5 the ground truth was obtained from manual segmentations. One center used the traditional approach in which different MRI sequences are combined for detecting calcification. In the other center a dedicated calcium MRI scan was made. Both methods resulted in reasonable volumes segmented for calcification, with automatic accuracies comparable to inter-observer variability. This suggests that MRI alone can be sufficient for calcium quantification, although CT remains more sensitive and accurate.

For lipid-rich necrotic core classification, slightly better correlations with the ground truth were obtained on the *in vivo* data compared to *ex vivo* MRI, which is probably owing to the use of contrast-enhanced MRI which better visualizes this component. In Chapter 5 we distinguished between lipid tissue and intraplaque hemorrhage, while in Chapters 2-3 any hemorrhage present is most likely to be included in the lipid-rich necrotic tissue. As the appearance of those components is quite different while LDC assumes each class to have Gaussian distribution, separating them, or using a more flexible classifier, may yield more accurate segmentation results. Moreover, intraplaque hemorrhage has shown to be a strong individual predictor of events [70, 160]. Ota et al. [141] showed that the agreement between MRI-determined hemorrhage and histology strongly depends on the imaging protocol used, with the MPRAGE scan as used in Chapter 5 obtaining the best accuracy. Indeed, in Chapter 5 we found a good agreement for the intraplaque-hemorrhage volumes for the data from both centers. That segmentation of hemorrhage in these protocols was mainly determined based on one image sequence, without the need to combine information from different images that need to be registered, is an advantage for automatic segmentation of this component.

We did not obtain accurate results for lipid-core segmentation when only using features derived from CTA in Chapter 3. This is in line with previous studies that had difficulties distinguishing fibrous and lipid tissue in CTA [45, 237]. However, it was recently shown that when considering the change in intensity between early and delayed enhancement (with an interval of 2 minutes), a much better relation with the amount of lipid or fibrous tissue was found than when using intensity in early enhancement [85]. This can be interesting to consider in future studies that compare imaging modalities.

7.2.2 Histology segmentation

We showed that automatic segmentation of histology sections can yield similar segmentation results in *ex vivo* and *in vivo* data as manual histology segmentation. There are a few things that need to be considered before automatic segmentation can be used to fully replace manual segmentation. We did not study the interobserver variability for histology segmentation. Moreover, it would be necessary to study how the results depend on which slices are selected for training by different observers, and the resulting variability in segmentation accuracy in the other imaging modalities.

In this thesis we studied training on different-specimen data and same-specimen data as two mutually exclusive approaches. An approach similar to the transfer-learning methods in Chapter 5 would be interesting to consider as well. A large set of samples from other specimens is available ('source data'), to which a few labelled samples from the new specimen can be added as 'target data' to train on. In addition, instead of accurately segmenting full sections manually, indicating a small number of pixels for each class that are representative for the entire specimen could yield accurate segmentations while requiring less user interaction. To take this one step further, a fully interactive approach could be considered, such as presented for foreground-background segmentation [155]. In our application more samples can be added iteratively until the segmentation is deemed sufficiently accurate. In Chapter 4 all histology sections were stained in one session using the same solutions for staining. When sections stained during different sessions are considered image appearance will be more variable, and the semi-automatic approach as presented in Chapter 4, or methods using transfer learning, are likely to be even more advantageous.

7.2.3 Ultrasound texture analysis

Chapter 6 showed the advantage of including carotid ultrasound texture measures for risk stratification. A number of improvements to this study can be considered. The current analysis combined all plaques from both vessels to obtain one risk parameter. It would be interesting to study if the single most vulnerable plaque can more accurately predict vascular risk, as an event in practice is caused by rupture of one plaque. However, it is difficult to determine which plaque has caused an event or is the most vulnerable. Previously the possibility of plaque component segmentation with respect to histology has been studied with moderate results [110]. With the inclusion of texture features better plaque component quantification may be obtained than when using intensity only, especially for the lipid or hemorrhagic tissue classes. In addition, contrast-enhanced ultrasound could enable the segmentation or quantification of neovascularization. However, it is unlikely that accuracy will be as high as with MRI because of a lower signal-to-noise ratio. In addition, acoustic shadowing hampers the analysis of parts of the plaque that are behind calcifications.

It should be noted that this study included only patients with known plaque and a stenosis <70% that were being followed in a stroke prevention clinic. Results may be more difficult to interpret with regard to initial development of atherosclerosis or highly advanced disease with stenosis degrees >70%. Currently medical treatment in the referred clinic is adjusted based on change in plaque volume, but based on our results it is likely advantageous to include plaque texture. To what extent the same accuracy can be expected on data acquired with a different ultrasound probe and/or different settings has not been studied. Images with a similar depth and resolution are generally similar in appearance. If necessary, separate models could be trained for data obtained with different settings or probes, to enable wide use in atherosclerosis monitoring.

In this thesis we did not study the predictive value of the obtained imaging parameters from MRI and CTA for (cerebro-)vascular events. The same approach as presented in Chapter 6 can be used to perform studies on risk assessment using a variety of parameters. The PARISk study [205, Chapter 5] aims to perform such an analysis in a longitudinal study in which MRI, CTA and ultrasound are obtained for a large set of symptomatic subjects with moderate carotid artery stenosis. Cox regression could show which parameters, obtained by which imaging modality, are most predictive and promising for use in daily clinical practice.

7.2.4 Segmentation methodology

Except for the methods used in Chapter 6, the presented segmentation methods are voxel classification algorithms. These algorithms have the advantage that complex models of local image appearance can easily be derived from the data. A drawback is that they may lead to noisy segmentations. However, the results we obtained were usually quite smooth with only few isolated voxels from different classes. This is probably related to the use of Gaussian filters at scales larger than the voxel size. One way to obtain more visually appealing segmentations would be to re-classify the few isolated voxels. Another option is to impose more spatial regularization using a graph-cut [211] or level-set approach [119], either directly on the images or on the posterior probabilities resulting from voxel classification. However, as said the segmentations are already relatively smooth and we expect that such changes in the methodology will have only a small effect on the estimated component volumes. The linear discriminant classifier that we used is one of the simplest and least flexible classifiers. Although the features used in this thesis generally did not fully satisfy the underlying assumptions of Gaussianity and equal class covariances, more flexible classifiers did not result in better segmentations in our experiments.

We have used both histology sections (Chapters 2-3) and manual annotations (Chapter 5) as ground truth for classifier training. Both have advantages and disadvantages. Histology sections enable very detailed and accurate assessment of plaque composition, which is an advantage when exact quantification and localization is desired. However, histology sectioning and staining is time consuming and subject to tissue deformations, both during excision and sectioning, which makes accurate registration with *in vivo* imaging challenging. Calcification, lipid and intraplaque hemorrhage visualization in different MRI sequences has extensively been validated with histology [159, 202, 243], and manual annotation can therefore be used as a ground truth. Manual annotation has the advantage that it does not require histology registration, and that it follows the image information better. This at the same time is a disadvantage since the exact quantification of components may be biased towards the imaging data, which for example results in the underestimation of lipid core size in MRI compared to histology [136]. In addition, due to noise and differences in interpretation variations between observers remain, as observed in Chapter 5. In such situations histology-guided annotations [119] would be beneficial as this can for example take away doubts on whether a certain area is calcified (dark in all sequences) or lipid (dark in postcontrast and T2w). Depending on the situation the appropriate ground truth may be defined. When presence of components, or a ranking of component volumes is sufficient, for example to perform clinical risk stratification, manual annotations suffice. When exact quantification is required, for example to determine which modality can most accurately visualize a certain component, histology sections would be preferred.

7.2.5 Remaining challenges for segmentation algorithms

We have shown that plaque component quantification in MRI, possibly in combination with CTA, and risk stratification from ultrasound is possible. In datasets for which a set of similar images with an adequate ground truth is available, image segmentation algorithms with good accuracy were developed. Results for volume quantification were good when comparing with histology and similar to interobserver variability. Most critical aspects in this respect seem image quality, imaging protocol, image normalization, and the need for a representative and sufficiently large set of training images in which all components are represented.

One issue with most current segmentation algorithms for carotid MRI is their limited ability to perform well on different data. Therefore the current challenge for image analysis lies mostly in the development of algorithms that are less sensitive to the specific dataset, scanner vendor, and protocol settings used. This is relevant in both clinical practice, and in research where larger and larger datasets are required with as a consequence more and larger multi-center studies. In CTA, owing to the Hounsfield scale, intensities across different scanners are similar, and algorithms for lumen segmentation were used on data from different scanners and scanner vendors [73]. In MRI there can be large variability between images acquired in different sessions, or with different scanners. In addition, the acquired image sequences in different carotid MRI studies varies greatly. A consensus on what images to acquire would greatly facilitate standardized quantification.

Another approach to overcome variability between MRI acquisition is quantitative T1- or T2-mapping in MRI. This provides absolute measurements of the T1 and T2 relaxation times, which are intrinsic physical tissue parameters. This has been applied in both *ex vivo* [46] and in *in vivo* MRI [18] plaque imaging. Major plaque components could be differentiated in these initial studies. If these quantitative MRI techniques are shown to be reproducible between centers and component segmentation has high accuracy this can stimulate wide implementation of plaque imaging and image analysis in clinical practice.

Although standardization of imaging protocols has a great advantage, it is unlikely that complete standardization will be reached. Differences between datasets, and between acquisitions will remain. In addition, due to advances in technology the state-of-the-art acquisition protocol will continuously change, and standardization is not possible in retrospective studies. In Chapter 5 we evaluated whether previously developed segmentation algorithms can be modified for use on new data when a limited amount of labels on the new data is obtained. This study showed some promising initial results and it is expected that research will continue in this direction. One interest for future research is to evaluate methods that one beforehand, based on the data, can indicate whether a certain approach will provide good results or not. In Chapter 5 whether adaptive histogram binning improved results or not, and the best performing transfer-learning method were different between the two centers. Several image similarity measures, or similarity of the probability density functions (PDF) of the vessel wall, could be evaluated. For the methods in Chapter 5 one requirement is still that there is a direct link between the features from the different datasets. In cases where this is not the case, as some features have a different ordering of classes, or the number of available images is different, different options can be used. The data acquisition parameters between studies could be compared to estimate which image sequences are most similar. A more explorative way could be to compare images based on the similarity of their PDF. A similar approach has been used to weight different-data training images based on the similarity of the PDF with the target image to be segmented [217]. If also the number of available images is different, it is possible to map different feature spaces to a shared space using for example domain adaptation metric learning [63].

7.2.6 Imaging modalities

In this thesis we started with the analysis of *ex vivo* MRI. Although *ex vivo* is not used in clinical practice, it can be useful in several situations. If accurate automatic segmentation is available it can be used in studies on plaque composition or plaque vulnerability, since much more detail can be seen than in *in vivo* MRI. Whereas *ex vivo* MRI is not as accurate as histology, its good correspondence to histology has been shown in multiple studies [38, 89, 170, Chapter 2] and could possibly replace labor-intensive histology processing and analysis in certain situations. It can also be used as an input for biomechanical modelling studies to assess the relation between plaque composition, tissue deformation and plaque stresses [88, 99]. Lastly, the presented results on importance of features and 3D registration can be used for translation to *in vivo* MRI image analysis, as we did in Chapter 3.

This thesis has furthermore focused on the non-invasive imaging modalities MRI, CTA and ultrasound. We showed that quantification of lipid-rich necrotic tissue, intraplaque hemorrhage, and calcification is possible from MRI, and that CTA can be used to more accurately quantify calcification. In addition, texture analysis from ultrasound can be used to stratify patients for risk for vascular events. These different imaging modalities are suitable for use at different stages of vascular disease.

Ultrasound is cheapest and most widely available and therefore most feasible for screening and monitoring of patients with (suspected) carotid artery disease. Screening of asymptomatic subjects with no risk factors is not recommended by clinical guidelines, but might be considered to detect atherosclerosis in subjects with two or more risk factors such as hypertension, hyperlipidemia, smoking, or family history of early atherosclerosis or ischemic stroke [26]. The presence of atherosclerosis, for example by means of IMT measurements, and stenosis degree by Doppler ultrasound, can then be determined. In case of plaque presence, ultrasound can be used for monitoring patients on a regular basis to determine whether a plaque is stable or progressing [26].

When presence of carotid artery disease has been confirmed by ultrasound, imaging by CTA and MRI becomes relevant in symptomatic patients or patients with considerable wall thickening. When intervention is considered, CTA or MRI can more accurately determine the stenosis degree, and show possible vascular lesions that are not seen on ultrasound, but are important for presurgical planning of intervention [26]. CTA imaging is faster and can accurately detect calcification, which is a contraindication for stenting, but not for CEA [17, 164] due to higher perioperative risk of complications. The advantages of MRI are a lack of ionizing radiation and more accurate depiction of plaque components.

Another imaging modality that has not been discussed yet, is Positron Emission Tomography (PET). As is discussed in more detail in the next section, PET imaging is most suitable for imaging inflammation and macrophage activity. These parameters are also considered important markers for plaque vulnerability [117]. Although PET alone would not be able to quantify intraplaque hemorrhage and lipid core, PET-MRI has become available as an experimental imaging modality [57]. Although these scanners are still very expensive, and only currently used in a few centers for experimental studies, atherosclerosis imaging would be a good application. The combination of PET and MRI would not only allow for imaging of calcification, lipid and intraplaque hemorrhage, but also for macrophage activity. However, due to high costs feasibility in clinical practice is doubtful until its superiority over especially PET/CT has been proven for several applications, of which oncology and neurologic disorders are most likely [79, 148].

7.2.7 Other imaging biomarkers

In addition to the plaque components studied in this thesis, there are other relevant plaque characteristics that relate to plaque vulnerability. One of the parameters that predict plaque rupture is a thin or ruptured fibrous cap [70]. In clinical MRI studies this is usually measured as a disrupted or discontinuous signal between the lumen and lipid core. However, as estimates for critical cap thickness are well below 1 mm [162], it is impossible to accurately measure the thickness of a thin cap with resolutions achievable with current *in vivo* imaging [136]. The segmentation results obtained in this thesis are therefore also not sufficient for accurate cap thickness measurements. In order for cap thickness to be used as a parameter in clinical risk assessment, or for segmentations to be used as input to biomechanical models to estimate cap stress, more accurate and smooth cap segmentations are required. One way to achieve this would be to use contour optimization instead of, or in addition to, voxel classification, to yield smoother segmentations. A method specifically designed to extract the fibrous cap is more likely to yield accurate results. One option would be to constrain the segmentation of a lipid core to be surrounded by fibrous tissue, like has been done for lumen and outer wall segmentation [209]. Parameters can then be optimized based on cap thickness accuracy. A different approach which may provide indirect information on cap thickness would be to look at intensity profiles from lumen to lipid core and relate those to cap thickness. However, these approaches would still be hampered by *in vivo* MRI resolution.

Plaque inflammation, measured by the presence of macrophages, is related to plaque development, progression and rupture [117]. Imaging of plaque inflammation has mostly been done with [^{18}F]-fluorodeoxyglucose (FDG) PET imaging [93]. Quantification of FDG uptake was shown to be highly sensitive and reproducible [156]. Moreover, inflammation as measured by PET was shown to decrease after high-dose statin use [191]. Limitations are the relatively low resolution of PET imaging and the need for a registered anatomical image such as CT or MRI.

Plaque inflammation has also been imaged with MRI. Contrast-enhanced MRI using an ultrasmall superparamagnetic iron oxide (USPIO) agent was shown to result in reduced signal intensity in T2*-weighted acquisitions at locations with macrophage presence [103, 203, 204]. Like with PET, also T2*-MRI has shown reduced macrophage activity after high-dose statin use [189]. MRI has the advantage that it is easier to combine with imaging of other plaque components in a multi-sequence MRI protocol, only at the expense of increased scanning time. A disadvantage of this technique is a delay of 24-48 hours between contrast administration and imaging.

Another feature of plaque vulnerability that received interest recently is intraplaque neovascularization. These neovessels within the plaque are leaky and therefore increase the amount of inflammatory cells and intraplaque hemorrhage presence [223]. Neovascularization can be visualized by contrast-enhanced ultrasound [196]. The contrast agent consists of microbubbles that can not leave the vasculature. Therefore microbubbles present in the plaque directly visualize intraplaque neovascularization. In addition, contrast-enhanced ultrasound enhances lumen visualization. Combination with B-mode ultrasound enables more accurate segmentation of the carotid artery lumen [5].

7.2.8 Applicability to other vessels

The carotid artery was the focus of this thesis, but atherosclerosis develops in other arteries as well. The other main arteries at risk are the coronary arteries, the aorta and the femoral artery in the lower extremities. Since disease in different vessel beds has been shown to be related, carotid imaging can be used as an indication for generalized atherosclerosis of the systemic circulation [90, 185]. In addition, image analysis tools developed in this thesis could be used in studies of other vessels.

For peripheral atherosclerosis an interest in MR imaging has recently developed [60, 78]. With a similar imaging protocol and resolution as for carotid imaging, images were obtained in which calcified and lipid-rich plaques could be distinguished [60]. It is likely that similar segmentation algorithms as developed for carotid MRI can be used for this data as well.

Due to the limited resolution MRI is currently not suited for studying vessel wall composition of the coronary arteries. For coronary artery imaging, CT is most popular, and methods for stenosis degree [100] and calcification quantification have been developed [169]. Additionally, the volume of non-calcified plaque as determined from coronary CTA and intravascular ultrasound (IVUS) was shown to be related, while the amount of low-density tissue in CTA was moderately correlated with fibro-fatty and necrotic tissue in IVUS [224]. The volume of non-calcified plaque was shown to be larger in patients with acute coronary syndrome [221]. Cardiac PET has mostly focused on imaging perfusion defects and ischemia. Although imaging of plaque inflammation is possible as well, studies on the coronary arteries have only had limited success due to limited resolution [11]. On the other hand, FDG-PET imaging of aorta was predictive of all cardiovascular events in a study on previously asymptomatic patients [54].

While we only studied the extracranial carotid bifurcation, intracranial carotid calcification was shown to be common in the elderly population [20]. It has been shown to be related to cognitive function, smaller brain tissue white matter volume and larger white matter lesion volumes [19, 21]. In addition, it was shown that lipid and intraplaque hemorrhage are more prevalent in symptomatic intracranial arteries [32] and therefore MR imaging of these arteries is of high interest as well [12].

7.2.9 Translation to clinical practice

Before image-based plaque characterization can be incorporated in clinical practice for diagnosis and to support clinical decision making, not only just accurate image analysis methods are required. Whether the use of imaging-derived measures leads to better patient care has to be proven as well. Several steps need to be taken in this process, of which Hlatky et al. [80] made an overview.

The first two requirements that are essential before taking further steps becomes relevant have already been met: the discrimination between subjects with and without outcome, and the predictive value of imaging-derived plaque composition have been shown in several studies for ultrasound [108, 109, Chapter 6], CT [104] and MRI [70, 160]. However, the incremental value in addition to established risk markers, and whether this sufficiently changes the predicted risk has only been validated up to some degree. The improved predictive value of ultrasound texture over the Framingham risk score [Chapter 6] and a set of features including plaque area, stenosis degree and history of symptoms [109] has been shown. For carotid MRI a number of studies performed univariate Cox regression and showed larger hazard ratios for measures of plaque composition than for other traditional clinical parameters [8, 163, 173, 188]. However, multivariate analysis was mostly not performed due to a limited number of events. Only one study showed better predictive value of intraplaque hemorrhage presence compared to traditional risk factors in a multivariate analysis [86].

In view of the above, well-designed and sufficiently large studies need to be performed to select the most suitable (combination of) imaging parameters, in addition to traditional risk parameters including stenosis degree and wall thickness. Possible imaging parameters are the presence or volume of intraplaque hemorrhage, lipid-rich necrotic tissue and/or calcification on MRI, calcification from CT, ultrasound texture, or parameters obtained by a combination of imaging modalities. Selection of parameters can be done in a study that acquires images in a large group of patients that will not be treated surgically but will be followed for events, such as PARISk [205] or CAIN [190]

Once these steps have been performed, it is time go ahead and study whether imaging-derived risk markers improve clinical outcome in a randomized clinical trial. In such a trial one group of patients is treated by current clinical guidelines, and for the other group decisions will be taken with the incorporation of measures of plaque composition. The formulation of a study criterion for treatment decision can then be based on the outcome of multivariate analysis in longitudinal studies as mentioned in the previous paragraph.

Finally, whether the use of new imaging modalities becomes the standard in clinical practice not only depends on the additional value to health care, but also on costs and feasibility. Cost-effectiveness studies are needed which balance any costs of additional imaging and interventions with the number of prevented events. Therefore, MRI must still have an advantage over risk stratification from ultrasound even if methodology such as presented in Chapter 6 is used, and use of PET-MRI will likely only become clinical practice if it proves to select patients for intervention with very high sensitivity and specificity.

7.3 Closing remarks

All in all, although there are several steps still to take, the work described in this thesis shows that automated characterization and quantification of carotid artery plaque composition from non-invasive imaging is possible. I hope that this work will be a good basis for further studies, and will thereby contribute to further refinement of methodology and future incorporation into clinical practice, with the final goal of improved patient care.

References

References

[1] V. Ablavsky, C. Becker, and P. Fua. Transfer learning by sharing support vectors. Technical report, School of Computer and Communication Sciences Swiss Federal Institute of Technology, Lausanne (EPFL), 2010.

[2] U. R. Acharya, S. V. Sree, M. M. R. Krishnan, F. Molinari, L. Saba, S. Y. S. Ho, A. T. Ahuja, S. C. Ho, A. Nicolaides, and J. S. Suri. Atherosclerotic risk stratification strategy for carotid arteries using texture-based features. *Ultrasound in Medicine Biology*, 38(6):899–915, 2012.

[3] I. M. Adame, R. J. van der Geest, B. A. Wasserman, M. A. Mohamed, J. H. C. Reiber, and B. P. F. Lelieveldt. Automatic segmentation and plaque characterization in atherosclerotic carotid artery MR images. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 16:227–234, 2004.

[4] C. D. Ainsworth, C. C. Blake, A. Tamayo, V. Beletsky, A. Fenster, and J. D. Spence. 3D ultrasound measurement of change in carotid plaque volume: A tool for rapid evaluation of new therapies. *Stroke*, 36(9):1904–1909, 2005.

[5] Z. Akkus, D. D. B. Carvalho, S. Klein, S. C. H. van den Oord, A. F. L. Schinkel, N. de Jong, A. F. W. van der Steen, and J. G. Bosch. Atherosclerotic carotid lumen segmentation in combined B-mode and contrast enhanced ultrasound images. In *Proc. of SPIE Medical Imaging*, 2014. in press.

[6] A. Akyildiz, L. Speelman, H. van Brummelen, M. Gutierrez, R. Virmani, A. van der Lugt, A. van der Steen, J. Wentzel, and F. Gijsen. Effects of intima stiffness and plaque morphology on peak cap stress. *BioMedical Engineering OnLine*, 10(1):25, 2011.

[7] L. Alic, J. C. Haeck, K. Bol, S. Klein, S. T. van Tiel, P. A. Wielepolski, M. de Jong, W. J. Niessen, M. Bernsen, and J. F. Veenland. Facilitating tumor functional assessment by spatially relating 3D tumor histology and *In Vivo* MRI: Image registration approach. *PLoS ONE*, 6(8):e22835, 2011.

[8] N. Altaf, L. Daniels, P. S. Morgan, D. Auer, S. T. MacSweeney, A. R. Moody, and J. R. Gladman. Detection of intraplaque hemorrhage by magnetic resonance imaging in symptomatic patients with mild to moderate carotid stenosis predicts recurrent neurological events. *Journal of Vascular Surgery*, 47(2):337–342, 2008.

[9] A. Alwan et al. *Global status report on noncommunicable diseases 2010.* World Health Organization, 2011.

[10] M. Amadasun and R. King. Textural features corresponding to textural properties. *Systems, Man and Cybernetics, IEEE Transactions on*, 19(5):1264–1274, 1989.

[11] B. M. Ardle, T. F. Dowsley, M. S. Cocker, H. Ohira, R. A. deKemp, J. DaSilva, T. D. Ruddy, B. J. Chow, and R. S. Beanlands. Cardiac PET: Metabolic and functional imaging of the myocardium. *Seminars in Nuclear Medicine*, 43(6):434–448, 2013.

[12] J. F. Arenillas. Intracranial atherosclerosis: Current concepts. *Stroke*, 42(1 suppl 1):S20–S23, 2011.

[13] A. Arias Lorza, D. D. B. Carvalho, J. Petersen, A. C. van Dijk, A. van der Lugt, W. J. Niessen, S. Klein, and M. de Bruijne. Carotid artery lumen segmentation in 3D free-hand ultrasound images using surface graph cuts. In *Medical image computing and computer assisted intervention*, 2013.

[14] A. Arias Lorza, J. Petersen, A. van Engelen, H. Tang, M. Selwaness, J. C. M. Witteman, A. van der Lugt, W. J. Niessen, and M. de Bruijne. Carotid artery wall segmentation by coupled surface graph cuts. In *MICCAI workshop on Medical Computer Vision, Nice*, 2012.

[15] Y. Artan, A. Oto, and I. Yetik. Cross-device automated prostate cancer localization with multiparametric MRI. *IEEE Transactions on Image Processing*, 22(12):5385–5394, 2013.

[16] J. Awad, A. Krasinski, G. Parraga, and A. Fenster. Texture analysis of carotid artery atherosclerosis from three-dimensional ultrasound images. *Medical Physics*, 37:1382–13991, 2010.

[17] E. R. Bates, J. D. Babb, D. E. Casey Jr, et al. ACCF/SCAI/SVMB/SIR/ASITN 2007 clinical expert consensus document on carotid stenting: A report of the american college of cardiology foundation task force on clinical expert consensus documents. *Journal of the American College of Cardiology*, 49(1):126–170, 2007.

[18] L. Biasiolli, A. C. Lindsay, J. T. Chai, R. P. Choudhury, and M. D. Robson. In-vivo quantitative T2 mapping of carotid arteries in atherosclerotic patients: segmentation and T2 measurement of plaque components. *Journal of Cardiovascular Magnetic Resonance*, 15:69, 2013.

[19] D. Bos, M. A. Ikram, S. E. Elias-Smale, G. P. Krestin, A. Hofman, J. C. M. Witteman, A. van der Lugt, and M. W. Vernooij. Calcification in major vessel beds relates to vascular brain disease. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 31(10):2331–2337, 2011.

[20] D. Bos, M. J. M. van der Rijk, T. E. A. Geeraedts, A. Hofman, G. P. Krestin, J. C. M. Witteman, A. van der Lugt, M. A. Ikram, and M. W. Vernooij. Intracranial carotid artery atherosclerosis: Prevalence and risk factors in the general population. *Stroke*, 43(7):1878–1884, 2012.

[21] D. Bos, M. W. Vernooij, S. E. Elias-Smale, B. F. J. Verhaaren, H. A. Vrooman, A. Hofman, W. J. Niessen, J. C. M. Witteman, A. van der Lugt, and M. A. Ikram. Atherosclerotic calcification relates to cognitive function and to brain changes on magnetic resonance imaging. *Alzheimer's & Dementia*, 8(5, Supplement):S10–S111, 2012.

[22] C. Bouveyron and S. Girard. Robust supervised classification with mixture models: Learning from data with uncertain labels. *Pattern Recognition*, 42(11):2649–2658, 2009.

[23] M. Brazzelli, K. Shuler, Z. Quayyum, D. Hadley, K. Muir, P. McNamee, J. De Wilde, M. Dennis, P. Sandercock, and J. M. Wardlaw. Clinical and imaging services for TIA and minor stroke: results of two surveys of practice across the UK. *BMJ Open*, 3(8), 2013.

[24] M. S. Breen, T. L. Lancaster, R. S. Lazebnik, S. G. Nour, J. S. Lewin, and D. L. Wilson. Three-dimensional method for comparing in vivo interventional MR images of thermally ablated tissue with tissue response. *Journal of Magnetic Resonance Imaging*, 18(1):90–102, 2003.

[25] D. J. Brenner and E. J. Hall. Computed tomography - an increasing source of radiation exposure. *New England Journal of Medicine*, 357(22):2277–2284, 2007.

[26] T. G. Brott, J. L. Halperin, S. Abbara, J. M. Bacharach, J. D. Barr, R. L. Bush, C. U. Cates, M. A. Creager, S. B. Fowler, G. Friday, V. S. Hertzberg, E. B. McIff, W. S. Moore, P. D. Panagos, T. S. Riles, R. H. Rosenwasser, and A. J. Taylor. 2011 ASA/ACCF/AHA/AANN/AANS/ACR/ASNR/CNS/SAIP/SCAI/SIR/SNIS/SVM/SVS guideline on the management of patients with extracranial carotid and vertebral artery disease: Executive summary. *Stroke*, 42(8):e420–e463, 2011.

[27] D. Buchanan, I. Gyacskov, E. Ukwatta, T. Lindenmaier, A. Fenster, and G. Parraga. Semi-automated segmentation of carotid artery total plaque volume from three dimensional ultrasound carotid imaging. In *Proc. of SPIE Vol. 8317*, 2012.

[28] J. Cai, T. S. Hatsukami, M. S. Ferguson, W. S. Kerwin, T. Saam, B. Chu, N. Takaya, N. L. Polissar, and C. Yuan. In vivo quantitative measurement of intact fibrous cap and lipid-rich necrotic core size in atherosclerotic carotid plaque: Comparison of high-resolution, contrast-enhanced magnetic resonance imaging and histology. *Circulation*, 112(22):3437–3444, 2005.

[29] J.-M. Cai, T. S. Hatsukami, M. S. Ferguson, R. Small, N. L. Polissar, and C. Yuan. Classification of human carotid atherosclerotic lesions with in vivo multicontrast magnetic resonance imaging. *Circulation*, 106(11):1368–1373, 2002.

[30] V. C. Cappendijk, S. Heeneman, A. G. H. Kessels, K. B. J. M. Cleutjens, G. W. H. Schurink, R. J. T. J. Welten, W. H. Mess, R.-J. van Suylen, T. Leiner, M. J. A. P. Daemen, J. M. A. van Engelshoven, and M. E. Kooi. Comparison of single-sequence T1w TFE MRI with multisequence MRI for the quantification of lipid-rich necrotic core in atherosclerotic plaque. *Journal of Magnetic Resonance Imaging*, 27(6):1347–1355, 2008.

[31] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[32] X. Y. Chen, K. S. Wong, W. W. M. Lam, H.-L. Zhao, and H. K. Ng. Middle cerebral artery atherosclerosis: Histological comparison between plaques associated with and not associated with infarct in a postmortem study. *Cerebrovascular Diseases*, 25:74–80, 2008.

[33] J. Cheng, H. Li, F. Xiao, A. Fenster, X. Zhang, X. He, L. Li, and M. Ding. Fully automatic plaque segmentation in 3-D carotid ultrasound images. *Ultrasound in Medicine Biology*, 39(12):2431–2446, 2013.

[34] B. Chiu, V. Shamdasani, R. Entrekin, C. Yuan, and W. S. Kerwin. Characterization of carotid plaques on 3-dimensional ultrasound imaging by registration with multicontrast magnetic resonance imaging. *Journal of Ultrasound in Medicine*, 31(10):1567–1580, 2012.

[35] C. I. Christodoulou, C. S. Pattichis, M. Pantziaris, and A. Nicolaides. Texture-based classification of atherosclerotic carotid plaques. *IEEE Transactions on Medical Imaging*, 22(7):902–912, 2003.

[36] B. Chu, A. Kampschulte, M. S. Ferguson, W. S. Kerwin, V. L. Yarnykh, K. D. O'Brien, N. L. Polissar, T. S. Hatsukami, and C. Yuan. Hemorrhage in the atherosclerotic carotid plaque: A high-resolution MRI study. *Stroke*, 35(5):1079–1084, 2004.

[37] S. E. Clarke, V. Beletsky, R. R. Hammond, R. A. Hegele, and B. K. Rutt. Validation of automatically classified magnetic resonance images for carotid plaque compositional analysis. *Stroke*, 37(1):93–97, 2006.

[38] S. E. Clarke, R. R. Hammond, J. R. Mitchell, and B. K. Rutt. Quantitative assessment of carotid plaque composition using multicontrast MRI and registered histology. *Magnetic Resonance in Medicine*, 50(6):1199–1208, 2003.

[39] C. A. Cocosco, A. P. Zijdenbos, and A. C. Evans. A fully automatic and robust brain MRI tissue classification method. *Medical Image Analysis*, 7(4):513–527, 2003.

[40] R. B. D'Agostino, R. S. Vasan, M. J. Pencina, P. A. Wolf, M. Cobain, J. M. Massaro, and W. B. Kannel. General cardiovascular risk profile for use in primary care: The Framingham heart study. *Circulation*, 117(6):743–753, 2008.

[41] J. Dauguet, T. Delzescaux, F. Condé, J.-F. Mangin, N. Ayache, P. Hantraye, and V. Frouin. Three-dimensional reconstruction of stained histological slices and 3D non-linear registration with in-vivo MRI for whole baboon brain. *Journal of Neuroscience Methods*, 164(1):191–204, 2007.

[42] M. H. Davis, A. Khotanzad, D. P. Flamig, and S. E. Harms. A physics-based coordinate transformation for 3-D image matching. *IEEE Transactions on Medical Imaging*, 16(3):317 –328, june 1997.

[43] C. de Monyé, F. Cademartiri, T. T. de Weert, D. A. M. Siepman, D. W. J. Dippel, and A. van Der Lugt. Sixteen-detector row CT angiography of carotid arteries: Comparison of different volumes of contrast material with and without a bolus chaser. *Radiology*, 237(2):555–562, 2005.

[44] C. de Monyé, T. T. de Weert, W. Zaalberg, F. Cademartiri, D. A. M. Siepman, D. W. J. Dippel, and A. van Der Lugt. Optimization of CT angiography of the carotid artery with a 16-MDCT scanner: Craniocaudal scan direction reduces contrast material-related perivenous artifacts. *American Journal of Roentgenology*, 186:1737–1745, 2006.

[45] T. T. de Weert, M. Ouhlous, E. Meijering, P. E. Zondervan, J. M. Hendriks, M. R. H. M. van Sambeek, D. W. J. Dippel, and A. van der Lugt. In vivo characterization and quantification of atherosclerotic carotid plaque components with multidetector computed tomography and histopathological correlation. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 26(10):2366–2372, 2006.

[46] A. J. Degnan, V. E. Young, T. Y. Tang, A. B. Gill, M. J. Graves, J. H. Gillard, and A. J. Patterson. Ex vivo study of carotid endarterectomy specimens: quantitative relaxation times within atherosclerotic plaque tissues. *Magnetic Resonance Imaging*, 30(7):1017–1021, 2012.

[47] H. M. den Ruijter, S. A. E. Peters, T. J. Anderson, A. R. Britton, J. M. Dekker, M. J. Eijkemans, G. Engström, G. W. Evans, J. de Graaf, D. E. Grobbee, et al. Common carotid intima-media thickness measurements in cardiovascular risk prediction: A meta-analysis. *JAMA*, 308(8):796–803, 2012.

[48] S. S. Dhawan, R. P. Nanjundappa, J. R. Branch, W. R. Taylor, A. A. Quyyumi, J. Hanjoong, M. C. McDaniel, J. Suo, D. Giddens, and H. Samady. Shear stress and plaque development. *Expert Review of Cardiovascular Therapy*, (4):545–556, 2010.

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

[50] R. P. W. Duin, P. Juszczak, P. Paclik, E. Pekalska, D. de Ridder, D. M. J. Tax, and S. Verzakov. *PRTools4.1, A Matlab Toolbox for Pattern Recognition*. Delft University of Technology, 2007. www.prtools.org.

[51] N. M. El-Barghouty, T. Levine, S. Ladva, A. Flanagan, and A. Nicolaides. Histological verification of computerised carotid plaque characterisation. *European Journal of Vascular and Endovascular Surgery*, 11(4):414–416, 1996.

[52] A. Fenster, D. B. Downey, and H. N. Cardinal. Three-dimensional ultrasound imaging. *Physics in Medicine and Biology*, 46(5):R67, 2001.

[53] A. Fenster, G. Parraga, and J. Bax. Three-dimensional ultrasound scanning. *Interface Focus*, 1(4):503–519, 2011.

[54] A. L. Figueroa, A. Abdelbaky, Q. A. Truong, E. Corsini, M. H. MacNabb, Z. R. Lavender, M. A. Lawler, S. K. Grinspoon, T. J. Brady, K. Nasir, U. Hoffmann, and A. Tawakol. Measurement of arterial activity on routine FDG PET/CT images improves prediction of risk of future CV events. *JACC: Cardiovascular Imaging*, 6(12):1250–1259, 2013.

[55] A. V. Finn, M. Nakano, J. Narula, F. D. Kolodgie, and R. Virmani. Concept of vulnerable/unstable plaque. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 30(7):1282–1292, 2010.

[56] B. Fischl, D. H. Salat, A. J. W. van der Kouwe, N. Makris, F. Ségonne, B. T. Quinn, and A. M. Dale. Sequence-independent segmentation of magnetic resonance images. *Neuroimage*, 23:S69–S84, 2004.

[57] F. Fraioli and S. Punwani. Clinical and research applications of simultaneous positron emission tomography and MRI. *British Journal of Radiology*, 87(1033), 2014.

[58] J. H. Friedman, T. Hastie, and R. Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22, 2 2010.

[59] M. H. Friedman, O. J. Deters, F. F. Mark, C. B. Bargeron, and G. M. Hutchins. Arterial geometry affects hemodynamics: A potential risk factor for atherosclerosis. *Atherosclerosis*, 46(2):225–231, 1983.

[60] M. S. Galizia, A. Barker, Y. Liao, J. Collins, J. Carr, M. M. McDermott, and M. Markl. Wall morphology, blood flow and wall shear stress: MR findings in patients with peripheral artery disease. *European Radiology*, pages 1–7, 2013.

[61] L. Gallardo Estrella, B. van Ginneken, and E. M. van Rikxoort. Normalization of CT scans reconstructed with different kernels to reduce variability in emphysema measurements. In *Proc. of SPIE Medical Imaging*, volume 8670, 2013.

[62] M. M. Galloway. Texture analysis using gray level run lengths. *Computer Graphics and Image Processing*, 4(2):172–179, 1975.

[63] B. Geng, D. Tao, and C. Xu. DAML: Domain adaptation metric learning. *IEEE Transactions on Image Processing*, 20(10):2980–2989, 2011.

[64] G. Geroulakos, G. Ramaswami, A. Nicolaides, K. James, N. Labropoulos, G. Belcaro, and M. Holloway. Characterization of symptomatic and asymptomatic carotid plaques using high-resolution real-time ultrasonography. *British Journal of Surgery*, 80(10):1274–1277, 1993.

[65] A. S. Go, D. Mozaffarian, V. L. Roger, et al. Heart disease and stroke statistics-2013 update: A report from the american heart association. *Circulation*, 127(1):e6–e245, 2013.

[66] T. Y. Goraya, S. J. Jacobsen, T. E. Kottke, R. L. Frye, S. A. Weston, and V. L. Roger. Coronary heart disease death and sudden cardiac death: A 20-year population-based study. *American Journal of Epidemiology*, 157(9):763–770, 2003.

[67] H. C. Groen, F. J. H. Gijsen, A. van der Lugt, M. S. Ferguson, T. S. Hatsukami, A. F. W. van der Steen, C. Yuan, and J. J. Wentzel. Plaque rupture in the carotid artery is localized at the high shear stress region: A case report. *Stroke*, 38(8):2379–2381, 2007.

[68] H. C. Groen, T. van Walsum, S. Rozie, S. Klein, K. van Gaalen, F. J. H. Gijsen, P. A. Wielopolski, H. M. M. van Beusekom, R. de Crom, H. J. M. Verhagen, A. F. W. van der Steen, A. van der Lugt, J. J. Wentzel, and W. J. Niessen. Three-dimensional registration of histology of human atherosclerotic carotid plaques to in-vivo imaging. *Journal of Biomechanics*, 43(11):2087–2092, 2010.

[69] M.-L. M. Grønholdt, B. G. Nordestgaard, T. V. Schroeder, S. Vorstrup, and H. Sillesen. Ultrasonic echolucent carotid plaques predict future strokes. *Circulation*, 104(1):68–73, 2001.

[70] A. Gupta, H. Baradaran, A. D. Schweitzer, H. Kamel, A. Pandya, D. Delgado, A. Dunning, A. I. Mushlin, and P. C. Sanelli. Carotid plaque MRI and stroke risk: A systematic review and meta-analysis. *Stroke*, 44(11):3071–3077, 2013.

[71] M. N. Gurcan, L. Boucheron, A. Can, A. Madabhushi, N. Rajpoot, and B. Yener. Histopathological image analysis: a review. *IEEE Reviews in Biomedical Engineering*, 2:147–171, 2009.

[72] K. Hameeteman, R. van 't Klooster, M. Selwaness, A. van der Lugt, J. C. M. Witteman, W. J. Niessen, and S. Klein. Carotid wall volume quantification from magnetic resonance images using deformable model fitting and learning-based correction of systematic errors. *Physics in Medicine and Biology*, 58(5):1605, 2013.

[73] K. Hameeteman, M. Zuluaga, M. Freiman, et al. Evaluation framework for carotid bifurcation lumen segmentation and stenosis grading. *Medical Image Analysis*, 15(4):477–488, 8 2011.

[74] K. Haraguchi, K. Houkin, I. Koyanagi, T. Nonaka, and T. Baba. Evaluation of carotid plaque composition by computed tomographic angiography and black blood magnetic resonance images. *Minimally Invasive Neurosurgery*, 51(2):91–94, 2008.

[75] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-3(6):610 –621, 1973.

[76] T. Hastie, R. Tibshirani, and J. H. Friedman. *The Elements of Statistical Learning, Chapter 4*. Springer, corrected edition, July 2003.

[77] T. S. Hatsukami, R. Ross, N. L. Polissar, and C. Yuan. Visualization of fibrous cap thickness and rupture in human atherosclerotic carotid plaque in vivo with high-resolution magnetic resonance imaging. *Circulation*, 102(9):959–964, 2000.

[78] K. Hayashi, V. Mani, A. Nemade, S. Aguiar, J. E. Postley, V. Fuster, and Z. A. Fayad. Variations in atherosclerosis and remodeling patterns in aorta and carotids. *Journal of Cardiovascular Magnetic Resonance*, 12(1):10, 2010.

[79] K. A. Herrmann, A. A. Kohan, M. C. Gaeta, C. Rubbert, J. L. Vercher-Conejero, R. M. Paspulati, K. Antonis, B. Mansoori, P. F. Faulhaber, N. Avril, and P. R. Ros. PET/MRI: Applications in clinical imaging. *Current Radiology Reports*, 1(3):161–176, 2013.

[80] M. A. Hlatky, P. Greenland, D. K. Arnett, C. M. Ballantyne, M. H. Criqui, M. S. Elkind, A. S. Go, F. E. Harrell, Y. Hong, B. V. Howard, V. J. Howard, P. Y. Hsue, C. M. Kramer, J. P. McConnell, S.-L. T. Normand, C. J. O'Donnell, S. C. Smith, P. W. Wilson, on behalf of the American Heart Association Expert Panel on Subclinical Atherosclerotic Diseases, E. R. Factors, and the Stroke Council. Criteria for evaluation of novel markers of cardiovascular risk: A scientific statement from the american heart association. *Circulation*, 119(17):2408–2416, 2009.

[81] V. Hodge and J. Austin. A survey of outlier detection methodologies. *Artificial Intelligence Review*, 22(2):85–126, 2004.

[82] J. M. A. Hofman, W. J. Branderhorst, H. M. M. ten Eikelder, V. C. Cappendijk, S. Heeneman, M. E. Kooi, P. A. J. Hilbers, and B. M. ter Haar Romeny. Quantification of atherosclerotic plaque components using in vivo MRI and supervised classifiers. *Magnetic Resonance in Medicine*, 55(4):790–799, 2006.

[83] M. Hollander, M. L. Bots, A. I. del Sol, P. J. Koudstaal, J. C. M. Witteman, D. E. Grobbee, A. Hofman, and M. M. B. Breteler. Carotid plaques increase the risk of stroke and subtypes of cerebral infarction in asymptomatic elderly. *Circulation*, 105(24):2872–2877, 2002.

[84] W. Hollingworth, A. B. Nathens, J. P. Kanne, M. L. Crandall, T. A. Crummy, D. K. Hallam, M. C. Wang, and J. G. Jarvik. The diagnostic accuracy of computed tomography angiography for traumatic or atherosclerotic lesions of the carotid and vertebral arteries: a systematic review. *European Journal of Radiology*, 48(1):88–102, 2003.

[85] N. Horie, M. Morikawa, S. Ishizaka, T. Takeshita, G. So, K. Hayashi, K. Suyama, and I. Nagata. Assessment of carotid plaque stability based on the dynamic enhancement pattern in plaque components with multidetector CT angiography. *Stroke*, 43(2):393–398, 2012.

[86] A. A. Hosseini, N. Kandiyil, S. T. S. MacSweeney, N. Altaf, and D. P. Auer. Carotid plaque hemorrhage on magnetic resonance imaging strongly predicts recurrent ischemia and stroke. *Annals of Neurology*, 73(6):774–784, 2013.

[87] G. N. Hounsfield. Computerized transverse axial scanning (tomography): Part 1. description of system. *British Journal of Radiology*, 46(552):1016–1022, 1973.

[88] X. Huang, C. Yang, J. Zheng, R. Bach, D. Muccigrosso, P. K. Woodard, and D. Tang. Higher critical plaque wall stress in patients who died of coronary artery disease compared with those who died of other causes: A 3D FSI study based on ex vivo MRI of coronary plaques. *Journal of Biomechanics*, 47(2):432–437, 2014.

[89] V. V. Itskovich, D. D. Samber, V. Mani, J. G. S. Aguinaldo, J. T. Fallon, C. Y. Tang, V. Fuster, and Z. A. Fayad. Quantification of human atherosclerotic plaques using spatially enhanced cluster analysis of multicontrast-weighted magnetic resonance images. *Magnetic Resonance in Medicine*, 52(3):515–523, 2004.

[90] T. Iwakiri, Y. Yano, Y. Sato, K. Hatakeyama, K. Marutsuka, S. Fujimoto, K. Kitamura, K. Kario, and Y. Asada. Usefulness of carotid intima-media thickness measurement as an indicator of generalized atherosclerosis: Findings from autopsy analysis. *Atherosclerosis*, 225(2):359–362, 2012.

[91] K. Jansen, A. F. W. van der Steen, H. M. M. van Beusekom, J. W. Oosterhuis, and G. van Soest. Intravascular photoacoustic imaging of human coronary atherosclerosis. *Optics Letters*, 36(5):597–599, 2011.

[92] K. Jansen, M. Wu, A. F. W. van der Steen, and G. van Soest. Lipid detection in atherosclerotic human coronaries by spectroscopic intravascular photoacoustic imaging. *Optics Express*, 21(18):21472–21484, 2013.

[93] F. Joshi, D. Rosenbaum, S. Bordes, and J. H. F. Rudd. Vascular imaging with positron emission tomography. *Journal of Internal Medicine*, 270(2):99–109, 2011.

[94] S. K. Kakkos, M. B. Griffin, A. N. Nicolaides, E. Kyriacou, M. M. Sabetai, T. Tegos, G. C. Makris, D. J. Thomas, and G. Geroulakos. The size of juxtaluminal hypoechoic area in ultrasound images of asymptomatic carotid plaques predicts the occurrence of stroke. *Journal of Vascular Surgery*, 57(3):609–618, 2013.

[95] M. K. Kalra, M. M. Maher, T. L. Toth, L. M. Hamberg, M. A. Blake, J.-A. Shepard, and S. Saini. Strategies for CT radiation dose optimization. *Radiology*, 230(3):619–628, 2004.

[96] R. K. Kanwar, R. Chaudhary, T. Tsuzuki, and J. R. Kanwar. Emerging engineered magnetic nanoparticulate probes for targeted MRI of atherosclerotic plaque macrophages. *Nanomedicine*, 7(5):735–749, 2012.

[97] C. Karmonik, P. Basto, K. Vickers, K. Martin, M. J. Reardon, G. M. Lawrie, and J. D. Morrisett. Quantitative segmentation of principal carotid atherosclerotic lesion components by feature space analysis based on multicontrast MRI at 1.5 T. *IEEE Transactions on Biomedical Engineering*, 56(2):352–360, 2009.

[98] W. Kerwin, D. Xu, F. Liu, T. Saam, H. Underhill, N. Takaya, B. Chu, T. Hatsukami, and C. Yuan. Magnetic resonance imaging of carotid atherosclerosis: Plaque analysis. *Topics in Magnetic Resonance Imaging*, 18(5):371–378, 2007.

[99] D. E. Kiousis, S. F. Rubinigg, M. Auer, and G. A. Holzapfel. A methodology to analyze changes in lipid core and calcification onto fibrous cap vulnerability: The human atherosclerotic carotid bifurcation as an illustratory example. *Journal of Biomechanical Engineering*, 131(12):121002, 2009.

[100] H. A. Kirişli, M. Schaap, C. T. Metz, et al. Standardized evaluation framework for evaluating coronary artery stenosis detection, stenosis quantification and lumen segmentation algorithms in computed tomography angiography. *Medical Image Analysis*, 17(8):859–876, 2013.

[101] S. Klein, J. Pluim, M. Staring, and M. Viergever. Adaptive stochastic gradient descent optimisation for image registration. *International Journal of Computer Vision*, 81:227–239, 2009.

[102] S. Klein, M. Staring, K. Murphy, M. Viergever, and J. Pluim. elastix: a toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, 29(1):196–205, 2010.

[103] M. E. Kooi, V. C. Cappendijk, K. B. J. M. Cleutjens, A. G. H. Kessels, P. J. E. H. M. Kitslaar, M. Borgers, P. M. Frederik, M. J. A. P. Daemen, and J. M. A. van Engelshoven. Accumulation of ultrasmall superparamagnetic particles of iron oxide in human atherosclerotic plaques can be detected by in vivo magnetic resonance imaging. *Circulation*, 107(19):2453–2458, 2003.

[104] R. M. Kwee. Systematic review on the association between calcification in carotid plaques and clinical ischemic symptoms. *Journal of Vascular Surgery*, 51(4):1015–1025, 2010.

[105] R. M. Kwee, M. T. B. Truijman, R. J. van Oostenbrugge, W. H. Mess, M. H. Prins, C. L. Franke, A. G. G. C. Korten, J. E. Wildberger, and M. E. Kooi. Longitudinal MRI study on the natural history of carotid artery plaques in symptomatic patients. *PLoS ONE*, 7(7):e42472, 07 2012.

[106] R. M. Kwee, J. M. A. van Engelshoven, W. H. Mess, J. W. M. ter Berg, F. H. B. M. Schreuder, C. L. Franke, A. G. G. C. Korten, B. J. Meems, R. J. van Oostenbrugge, J. E. Wildberger, and M. E. Kooi. Reproducibility of fibrous cap status assessment of carotid artery plaques by contrast-enhanced MRI. *Stroke*, 40(9):3017–3021, 2009.

[107] R. M. Kwee, R. J. van Oostenbrugge, L. Hofstra, G. J. Teule, J. van Engelshoven, W. H. Mess, and M. E. Kooi. Identifying vulnerable carotid plaques by noninvasive imaging. *Neurology*, 70(24 Part 2):2401–2409, 2008.

[108] E. C. Kyriacou, C. Pattichis, M. Pattichis, C. Loizou, C. Christodoulou, S. K. Kakkos, and A. Nicolaides. A review of noninvasive ultrasound image processing methods in the analysis of carotid plaque morphology for the assessment of stroke risk. *IEEE Transactions on Information Technology in Biomedicine*, 14(4):1027–1038, 2010.

[109] E. C. Kyriacou, S. Petroudi, C. S. Pattichis, M. S. Pattichis, M. Griffin, S. Kakkos, and A. Nicolaides. Prediction of high-risk asymptomatic carotid plaques based on ultrasonic image features. *IEEE Transactions on Information Technology in Biomedicine*, 16(5):966 –973, 2012.

[110] B. K. Lal, R. W. Hobson II, M. Hameed, P. J. Pappas, F. T. Padberg Jr., Z. Jamil, and W. N. Durán. Noninvasive identification of the unstable carotid plaque. *Annals of Vascular Surgery*, 20(2):167–174, 2006.

[111] A. Landry, J. D. Spence, and A. Fenster. Measurement of carotid plaque volume by 3-dimensional ultrasound. *Stroke*, 35(4):864–869, 2004.

[112] P. Lauterbur. Image formation by induced local interactions: Examples employing nuclear magnetic resonance. *Nature*, 242:190–191, 1973.

[113] K. I. Laws. Rapid texture identification. In T. F. Wiener, editor, *Image Processing for Missile Guidance. SPIE Proceedings on*, volume 238, pages 376–381, 1980.

[114] C. L. Lendon, M. J. Davies, G. V. R. Born, and P. D. Richardson. Atherosclerotic plaque caps are locally weakened when macrophages density is increased. *Atherosclerosis*, 87(1):87–90, 1991.

[115] F. Li, V. L. Yarnykh, T. S. Hatsukami, B. Chu, N. Balu, J. Wang, H. R. Underhill, X. Zhao, R. Smith, and C. Yuan. Scan-rescan reproducibility of carotid atherosclerotic plaque morphology and tissue composition measurements using multicontrast MRI at 3T. *Journal of Magnetic Resonance Imaging*, 31(1):168–176, 2010.

[116] C. D. Liapis, J. D. Kakisis, and A. G. Kostakis. Carotid stenosis: Factors affecting symptomatology. *Stroke*, 32(12):2782–2786, 2001.

[117] P. Libby. Inflammation in atherosclerosis. *Nature*, 420:868–874, 2002.

[118] X. Lin and X.-W. Chen. Mr.KNN: soft relevance for multi-label classification. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, CIKM '10, pages 349–358. ACM, 2010.

[119] F. Liu, D. Xu, M. S. Ferguson, B. Chu, T. Saam, N. Takaya, T. S. Hatsukami, C. Yuan, and W. S. Kerwin. Automated in vivo segmentation of carotid plaque MRI with morphology-enhanced probability maps. *Magnetic Resonance in Medicine*, 55(3):659–668, 2006.

[120] A. Long, A. Lepoutre, Corbillon, and A. Branchereau. Critical review of non- or minimally invasive methods (duplex ultrasonography, MR- and CT-angiography) for evaluating stenosis of the proximal internal carotid artery. *European Journal of Vascular and Endovascular Surgery*, 24(1):43–52, 2002.

[121] M. W. Lorenz, H. S. Markus, M. L. Bots, M. Rosvall, and M. Sitzer. Prediction of clinical cardiovascular events with carotid intima-media thickness: A systematic review and meta-analysis. *Circulation*, 115(4):459–467, 2007.

[122] M. W. Lorenz, J. F. Polak, M. Kavousi, et al. Carotid intima-media thickness progression to predict cardiovascular events in the general population (the PROG-IMT collaborative project): a meta-analysis of individual participant data. *The Lancet*, 379(9831):2053–2062, 2012.

[123] J. K. Lovett, P. J. Gallagher, and P. M. Rothwell. Reproducibility of histological assessment of carotid plaque: Implications for studies of carotid imaging. *Cerebrovascular Diseases*, 18:117–123, 2004.

[124] J. K. Lovett, J. N. E. Redgrave, and P. M. Rothwell. A critical appraisal of the performance, reporting, and interpretation of studies comparing carotid plaque imaging with histology. *Stroke*, 36(5):1085–1091, 2005.

[125] M. R. Makowski and R. M. Botnar. MR imaging of the arterial vessel wall: Molecular imaging from bench to bedside. *Radiology*, 269(1):34–51, 2013.

[126] R. Manniesing, M. Schaap, S. Rozie, K. Hameeteman, D. Vukadinovic, A. van der Lugt, and W. J. Niessen. Robust CTA lumen segmentation of the atherosclerotic carotid artery bifurcation in a large patient population. *Medical Image Analysis*, 14(6):759–69, 12 2010.

[127] J. McNamara, M. A. Molot, J. F. Stremple, and R. T. Cutting. Coronary artery disease in combat casualties in vietnam. *JAMA*, 216(7):1185–1187, 1971.

[128] S. Mendis, P. Puska, B. Norrving, et al. *Global atlas on cardiovascular disease prevention and control.* World Health Organization, 2011.

[129] L. M. Mitsumori, T. S. Hatsukami, M. S. Ferguson, W. S. Kerwin, J. Cai, and C. Yuan. In vivo accuracy of multisequence MR imaging for identifying unstable fibrous caps in advanced human carotid plaques. *Journal of Magnetic Resonance Imaging*, 17(4):410–420, 2003.

[130] J. Morrisett, W. Vick, R. Sharma, G. Lawrie, M. Reardon, E. Ezell, J. Schwartz, G. Hunter, and D. Gorenstein. Discrimination of components in atherosclerotic plaques from human carotid endarterectomy specimens by magnetic resonance imaging ex vivo. *Magnetic Resonance Imaging*, 21(5):465–474, 2003.

[131] M. M. Mughal, M. K. Khan, J. K. DeMarco, A. Majid, F. Shamoun, and G. S. Abela. Symptomatic and asymptomatic carotid artery plaque. *Expert Review of Cardiovascular Therapy*, 9(10):1315–1330, 2011.

[132] G. Mühlenbruch, M. Das, G. Mommertz, M. Schaaf, S. Langer, A. Mahnken, J. Wildberger, A. Thron, R. Günther, and T. Krings. Comparison of dual-source CT angiography and MR angiography in preoperative evaluation of intra- and extracranial vessels: a pilot study. *European Radiology*, 20(2):469–476, 2010.

[133] M. Naghavi, P. Libby, E. Falk, et al. From vulnerable plaque to vulnerable patient: A call for new definitions and risk assessment strategies: Part I. *Circulation*, 108(14):1664–1672, 2003.

[134] A. Nair, B. D. Kuban, E. M. Tuzcu, P. Schoenhagen, S. E. Nissen, and D. G. Vince. Coronary plaque classification with intravascular ultrasound radiofrequency data analysis. *Circulation*, 106(17):2200–2206, 2002.

[135] A. N. Nicolaides, S. K. Kakkos, E. Kyriacou, M. Griffin, M. Sabetai, D. J. Thomas, T. Tegos, G. Geroulakos, N. Labropoulos, C. J. Doré, T. P. Morris, R. Naylor, and A. L. Abbott. Asymptomatic internal carotid artery stenosis and cerebrovascular risk stratification. *Journal of Vascular Surgery*, 52(6):1486–1496, 2010.

[136] H. A. Nieuwstadt, T. R. Geraedts, M. T. B. Truijman, M. E. Kooi, A. van der Lugt, A. F. W. van der Steen, J. J. Wentzel, M. Breeuwer, and F. J. H. Gijsen. Numerical simulations of carotid MRI quantify the accuracy in measuring atherosclerotic plaque components in vivo. *Magnetic Resonance in Medicine*, 2013. epub ahead of print.

[137] S. E. Nissen, J. C. Gurley, C. L. Grines, D. C. Booth, R. McClure, M. Berk, C. Fischer, and A. N. DeMaria. Intravascular ultrasound assessment of lumen size and wall morphology in normal subjects and patients with coronary artery disease. *Circulation*, 84(3):1087–99, 1991.

[138] L. G. Nyul, J. K. Udupa, and X. Zhang. New variants of a method of MRI scale standardization. *IEEE Transactions on Medical Imaging*, 19(2):143–150, 2000.

[139] T. Ojala, M. Pietikäinen, and T. Mäenpää. Gray scale and rotation invariant texture classification with local binary patterns. In *Computer Vision - ECCV 2000*, volume 1842 of *Lecture Notes in Computer Science*, pages 404–420. Springer Berlin / Heidelberg, 2000.

[140] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971 –987, 2002.

[141] H. Ota, V. L. Yarnykh, M. S. Ferguson, H. R. Underhill, J. K. DeMarco, D. C. Zhu, M. Oikawa, L. Dong, X. Zhao, A. Collar, T. S. Hatsukami, and C. Yuan. Carotid intraplaque hemorrhage imaging at 3.0-T MR imaging: Comparison of the diagnostic performance of three T1-weighted sequences. *Radiology*, 254(2):551–563, 2010.

[142] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010.

[143] K. I. Paraskevas, D. P. Mikhailidis, and F. J. Veith. Comparison of the five 2011 guidelines for the treatment of carotid stenosis. *Journal of Vascular Surgery*, 55(5):1504–1508, 2012.

[144] K. I. Paraskevas, F. J. Veith, T. S. Riles, and W. S. Moore. Is carotid artery stenting a fair alternative to carotid endarterectomy for symptomatic carotid artery stenosis? a commentary on the AHA/ASA guidelines. *Journal of Vascular Surgery*, 54(2):541–543, 2011.

[145] A. E. Park, W. J. McCarthy, W. H. Pearce, J. S. Matsumura, and J. S. Yao. Carotid plaque morphology correlates with presenting symptomatology. *Journal of Vascular Surgery*, 27(5):872–879, 1998.

[146] G. W. Petty, J. Brown, Robert D., J. P. Whisnant, J. D. Sicks, W. M. O'Fallon, and D. O. Wiebers. Ischemic stroke subtypes: A population-based study of incidence and risk factors. *Stroke*, 30(12):2513–2516, 1999.

[147] R. H. H. M. Philipsen, P. Maduskar, L. Hogeweg, and B. van Ginneken. Normalization of chest radiographs. In *Proc. of SPIE Medical Imaging*, volume 8670, 2013.

[148] B. J. Pichler, A. Kolb, T. Nägele, and H.-P. Schlemmer. PET/MRI: Paving the way for the next generation of clinical multimodality imaging applications. *Journal of Nuclear Medicine*, 51(3):333–336, 2010.

[149] J. F. Polak, L. Shemanski, D. H. O'Leary, D. Lefkowitz, T. R. Price, P. J. Savage, W. E. Brant, and C. Reid. Hypoechoic plaque at US of the carotid artery: an independent risk factor for incident stroke in adults aged 65 years or older. cardiovascular health study. *Radiology*, 208(3):649–654, 1998.

[150] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.

[151] S. P. Raman, P. T. Johnson, S. Deshmukh, M. Mahesh, K. L. Grant, and E. K. Fishman. CT dose reduction applications: Available tools on the latest generation of CT scanners. *Journal of the American College of Radiology*, 10(1):37–41, 2013.

[152] J. J. Ricotta, A. AbuRahma, E. Ascher, M. Eskandari, P. Faries, and B. K. Lal. Updated society for vascular surgery guidelines for management of extracranial carotid disease. *Journal of Vascular Surgery*, 54(3):e1–e31, 2011.

[153] K. Rohr, H. S. Stiehl, R. Sprengel, T. M. Buzug, J. Weese, and M. H. Kuhn. Landmark-based elastic registration using approximating thin-plate splines. *IEEE Transactions on Medical Imaging*, 20(6):526–534, 2001.

[154] R. R. Ronen, S. E. Clarke, R. R. Hammond, and B. K. Rutt. Carotid plaque classification: Defining the certainty with which plaque components can be differentiated. *Magnetic Resonance in Medicine*, 57(5):874–880, 2007.

[155] C. Rother, V. Kolmogorov, and A. Blake. "GrabCut": Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23(3):309–314, Aug. 2004.

[156] J. H. F. Rudd, K. S. Myers, S. Bansilal, J. Machac, C. A. Pinto, C. Tong, A. Rafique, R. Hargeaves, M. Farkouh, V. Fuster, and Z. A. Fayad. Atherosclerosis inflammation imaging with 18F-FDG PET: Carotid, iliac, and femoral uptake reproducibility, quantification methods, and recommendations. *Journal of Nuclear Medicine*, 49(6):871–878, 2008.

[157] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999.

[158] T. Saam, M. S. Ferguson, V. L. Yarnykh, N. Takaya, D. Xu, N. L. Polissar, T. S. Hatsukami, and C. Yuan. Quantitative evaluation of carotid plaque composition by in vivo MRI. *Arterioscler Thromb Vasc Biol*, 25(1):234–239, 2005.

[159] T. Saam, T. S. Hatsukami, N. Takaya, B. Chu, H. Underhill, W. S. Kerwin, J. Cai, M. S. Ferguson, and C. Yuan. The vulnerable, or high-risk, atherosclerotic plaque: Noninvasive MR imaging for characterization and assessment. *Radiology*, 244(1):64–77, 2007.

[160] T. Saam, H. Hetterich, V. Hoffmann, C. Yuan, M. Dichgans, H. Poppert, T. Koeppel, U. Hofmann, M. F. Reiser, and F. Bamberg. Meta-analysis and systematic review of the predictive value of carotid plaque hemorrhage on cerebrovascular events by magnetic resonance imaging. *Journal of the American College of Cardiology*, 62(12):1081–1091, 2013.

[161] L. Saba, M. Anzidei, B. C. Marincola, M. Piga, E. Raz, P. P. Bassareo, A. Napoli, L. Mannelli, C. Catalano, and M. Wintermark. Imaging of the carotid artery vulnerable plaque. *CardioVascular and Interventional Radiology*, pages 1–14, 2013.

[162] L. Saba, F. Potters, A. Lugt, and G. Mallarini. Imaging of the fibrous cap in atherosclerotic carotid plaque. *CardioVascular and Interventional Radiology*, 33(4):681–689, 2010.

[163] U. Sadat, Z. Teng, V. E. Young, S. Walsh, Z. Y. Li, M. J. Graves, K. Varty, and J. H. Gillard. Association between biomechanical structural stresses of atherosclerotic carotid plaques and subsequent ischaemic cerebrovascular events - a longitudinal in vivo magnetic resonance imaging-based finite element study. *European Journal of Vascular and Endovascular Surgery*, 40(4):485–491, 2010.

[164] R. D. Safian. Treatment strategies for carotid stenosis in patients at increased risk for surgery. *Progress in Cardiovascular Diseases*, 54(1):22–28, 2011.

[165] M. A. Saleem, U. Sadat, S. R. Walsh, V. E. Young, J. H. Gillard, D. G. Cooper, and M. E. Gaunt. Role of carotid duplex imaging in carotid screening programmes–an overview. *Cardiovascular ultrasound*, 6(1):34, 2008.

[166] J. Sanz and Z. A. Fayad. Imaging of atherosclerotic cardiovascular disease. *Nature*, 451(7181):953–957, 2008.

[167] J. M. Seeger, E. Barratt, G. A. Lawson, and N. Klingman. The relationship between carotid plaque composition, plaque morphology, and neurologic symptoms. *Journal of Surgical Research*, 58(3):330–336, 1995.

[168] W. E. Shaalan, H. Cheng, B. Gewertz, J. F. McKinsey, L. B. Schwartz, D. Katz, D. Cao, T. Desai, S. Glagov, and H. S. Bassiouny. Degree of carotid plaque calcification in relation to symptomatic outcome and plaque inflammation. *Journal of Vascular Surgery*, 40(2):262–269, 2004.

[169] R. Shahzad, T. van Walsum, M. Schaap, A. Rossi, S. Klein, A. C. Weustink, P. J. de Feyter, L. J. van Vliet, and W. J. Niessen. Vessel specific coronary artery calcium scoring: An automatic system. *Academic Radiology*, 20(1):1–9, 2013.

[170] M. Shinnar, J. T. Fallon, S. Wehrli, M. Levin, D. Dalmacy, Z. A. Fayad, J. J. Badimon, M. Harrington, E. Harrington, and V. Fuster. The diagnostic accuracy of ex vivo MRI for human atherosclerotic plaque characterization. *Arterioscler Thromb Vasc Biol*, 19(11):2756–2761, 1999.

[171] H. Sillesen and E. Falk. Why not screen for subclinical atherosclerosis? *The Lancet*, 378(9792):645–646, 2011.

[172] N. Simon, J. H. Friedman, T. Hastie, and R. Tibshirani. Regularization paths for Cox's proportional hazards model via coordinate descent. *Journal of Statistical Software*, 39(5):1–13, 3 2011.

[173] N. Singh, A. R. Moody, D. J. Gladstone, G. Leung, R. Ravikumar, J. Zhan, and R. Maggisano. Moderate carotid artery stenosis: MR imaging-depicted intraplaque hemorrhage predicts risk of cerebrovascular ischemic events in asymptomatic men. *Radiology*, 252(2):502–508, 2009.

[174] C. J. Slager, J. J. Wentzel, F. J. H. Gijsen, J. C. H. Schuurbiers, A. C. van der Wal, A. F. W. van der Steen, and P. W. Serruys. The role of shear stress in the generation of rupture-prone vulnerable plaques. *Nature Reviews Cardiology*, 2(8):401–407, 2005.

[175] C. J. Slager, J. J. Wentzel, F. J. H. Gijsen, A. Thury, A. C. van der Wal, J. A. Schaar, and P. W. Serruys. The role of shear stress in the destabilization of vulnerable plaques and related therapeutic implications. *Nature Reviews Cardiology*, 2(9):456–464, 2005.

[176] J. G. Sled, A. P. Zijdenbos, and A. C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical Imaging*, 17(1):87–97, 1998.

[177] Ö. Smedby and L. Bergstrand. Tortuosity and atherosclerosis in the femoral artery: What is cause and what is effect? *Annals of Biomedical Engineering*, 24(4):474–480, 1996.

[178] L.-K. Soh and C. Tsatsoulis. Texture analysis of sar sea ice imagery using gray level co-occurrence matrices. *IEEE Transactions on Geoscience and Remote Sensing*, 37(2):780 –795, 1999.

[179] L. Sorensen, S. B. Shaker, and M. de Bruijne. Quantitative analysis of pulmonary emphysema using local binary patterns. *IEEE Transactions on Medical Imaging*, 29(2):559 –569, 2010.

[180] R. M. Sousa, C. P. Ferri, D. Acosta, E. Albanese, M. Guerra, Y. Huang, K. S. Jacob, A. T. Jotheeswaran, J. J. L. Rodriguez, G. R. Pichardo, M. C. Rodriguez, A. Salas, A. L. Sosa, J. Williams, T. Zuniga, and M. Prince. Contribution of chronic diseases to disability in elderly people in countries with low and middle incomes: a 10/66 dementia research group population-based survey. *The Lancet*, 374(9704):1821 – 1830, 2009.

[181] J. D. Spence, V. Coates, H. Li, A. Tamayo, C. Munoz, D. Hackam, M. DiCicco, J. DesRoches, C. Bogiatzi, J. Klein, J. Madrenas, and R. A. Hegela. Effects of intensive medical therapy on microemboli and cardiovascular risk in asymptomatic carotid stenosis. *Archives of Neurology*, 67(2):180–186, 2010.

[182] J. D. Spence, M. Eliasziw, M. DiCicco, D. G. Hackam, R. Galil, and T. Lohmann. Carotid plaque area: A tool for targeting and evaluating vascular preventive therapy. *Stroke*, 33(12):2916–2922, 2002.

[183] J. D. Spence and D. G. Hackam. Treating arteries instead of risk factors: A paradigm change in management of atherosclerosis. *Stroke*, 41(6):1193–1199, 2010.

[184] H. C. Stary. Natural history and histological classification of atherosclerotic lesions: An update. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 20(5):1177–1178, 2000.

[185] J. H. Stein, C. E. Korcarz, R. T. Hurst, E. Lonn, C. B. Kendall, E. R. Mohler, S. S. Najjar, C. M. Rembold, and W. S. Post. Use of carotid ultrasound to identify subclinical vascular disease and evaluate cardiovascular disease risk: A consensus statement from the american society of echocardiography carotid intima-media thickness task force endorsed by the society for vascular medicine. *Journal of the American Society of Echocardiography*, 21(2):93–111, 2008.

[186] J. Sun, N. Balu, D. S. Hippe, Y. Xue, L. Dong, X. Zhao, F. Li, D. Xu, T. S. Hatsukami, and C. Yuan. Subclinical carotid atherosclerosis: Short-term natural history of lipid-rich necrotic core - a multicenter study with MR imaging. *Radiology*, 268(1):61–68, 2013.

[187] N. Takaya, J. Cai, M. S. Ferguson, V. L. Yarnykh, B. Chu, T. Saam, N. L. Polissar, J. Sherwood, R. C. Cury, R. J. Anders, K. O. Broschat, D. Hinton, K. L. Furie, T. S. Hatsukami, and C. Yuan. Intra- and interreader reproducibility of magnetic resonance imaging for quantifying the lipid-rich necrotic core is improved with gadolinium contrast enhancement. *Journal of Magnetic Resonance Imaging*, 24(1):203–210, 2006.

[188] N. Takaya, C. Yuan, B. Chu, T. Saam, H. Underhill, J. Cai, N. Tran, N. L. Polissar, C. Isaac, M. S. Ferguson, G. A. Garden, S. C. Cramer, K. R. Maravilla, B. Hashimoto, and T. S. Hatsukami. Association between carotid plaque characteristics and subsequent ischemic cerebrovascular events: A prospective assessment with MRI - initial results. *Stroke*, 37(3):818–823, 2006.

[189] T. Y. Tang, S. P. S. Howarth, S. R. Miller, et al. The ATHEROMA (atorvastatin therapy: Effects on reduction of macrophage activity) study: Evaluation using ultrasmall superparamagnetic iron oxide-enhanced magnetic resonance imaging in carotid disease. *Journal of the American College of Cardiology*, 53(22):2039–2050, 2009.

[190] J.-C. Tardif, J. D. Spence, T. M. Heinonen, A. Moody, J. Pressacco, R. Frayne, P. L'Allier, B. J. W. Chow, M. Friedrich, S. E. Black, A. Fenster, B. Rutt, and R. Beanlands. Atherosclerosis imaging and the Canadian atherosclerosis imaging network. *Canadian Journal of Cardiology*, 29(3):297–303, 2013.

[191] A. Tawakol, Z. A. Fayad, R. Mogg, A. Alon, M. T. Klimas, H. Dansky, S. S. Subramanian, A. Abdelbaky, J. H. F. Rudd, M. E. Farkouh, I. O. Nunes, C. R. Beals, and S. S. Shankar. Intensification of statin therapy results in a rapid reduction in atherosclerotic inflammation: Results of a multicenter fluorodeoxyglucose-positron emission tomography/computed tomography feasibility study. *Journal of the American College of Cardiology*, 62(10):909–917, 2013.

[192] D. M. J. Tax. DDtools, the data description toolbox for Matlab, May 2012. version 1.9.1, http://prlab.tudelft.nl/david-tax/dd_tools.html.

[193] B. C. te Boekhorst, G. A. van Tilborg, G. J. Strijkers, and K. Nicolay. Molecular MRI of inflammation in atherosclerosis. *Curr Cardiovasc Imaging Rep*, 5:60–68, 2012.

[194] G. J. Tearney, E. Regar, T. Akasaka, et al. Consensus standards for acquisition, measurement, and reporting of intravascular optical coherence tomography studies: A report from the international working group for intravascular optical coherence tomography standardization and validation. *Journal of the American College of Cardiology*, 59(12):1058–1072, 2012.

[195] G. L. ten Kate, E. J. Sijbrands, D. Staub, B. Coll, F. J. ten Cate, S. B. Feinstein, and A. F. L. Schinkel. Noninvasive imaging of the vulnerable atherosclerotic plaque. *Current Problems in Cardiology*, 35(11):556–591, 2010.

[196] G. L. ten Kate, S. C. H. van den Oord, E. J. G. Sijbrands, A. van der Lugt, N. de Jong, J. G. Bosch, A. F. W. van der Steen, and A. F. L. Schinkel. Current status and future developments of contrast-enhanced ultrasound of carotid atherosclerosis. *Journal of Vascular Surgery*, 57(2):539–546, 2013.

[197] M. Tendera, V. Aboyans, M.-L. Bartelink, et al. ESC guidelines on the diagnosis and treatment of peripheral artery diseases: Document covering atherosclerotic disease of extracranial carotid and vertebral, mesenteric,

renal, upper and lower extremity arteriesthe task force on the diagnosis and treatment of peripheral artery diseases of the european society of cardiology (ESC). *European Heart Journal*, 32(22):2851–2906, 2011.

[198] The European Stroke Organisation (ESO) Executive Committee and the ESO Writing Committee. Guidelines for management of ischaemic stroke and transient ischaemic attack 2008. *Cerebrovascular Diseases*, 25:457–507, 2008.

[199] P. Thevenaz and M. Unser. Optimization of mutual information for multiresolution image registration. *IEEE Transactions on Image Processing*, 9(12):2083–2099, 2000.

[200] C. Thiel. Classification on soft labels is robust against label noise. In I. Lovrek, R. Howlett, and L. Jain, editors, *Knowledge-Based Intelligent Information and Engineering Systems*, volume 5177 of *Lecture Notes in Computer Science*, pages 65–73. Springer Berlin / Heidelberg, 2008.

[201] J. M. Tobis, J. Mallery, D. Mahon, K. Lehmann, P. Zalesky, J. Griffith, J. Gessert, M. Moriuchi, M. McRae, and M. L. Dwyer. Intravascular ultrasound imaging of human coronary arteries in vivo. analysis of tissue characterizations with comparison to in vitro histological specimens. *Circulation*, 83(3):913–26, 1991.

[202] J.-F. Toussaint, G. M. LaMuraglia, J. F. Southern, V. Fuster, and H. L. Kantor. Magnetic resonance images lipid, fibrous, calcified, hemorrhagic, and thrombotic components of human atherosclerosis in vivo. *Circulation*, 94(5):932–938, 1996.

[203] R. A. Trivedi, C. Mallawarachi, J.-M. U-King-Im, M. J. Graves, J. Horsley, M. J. Goddard, A. Brown, L. Wang, P. J. Kirkpatrick, J. Brown, and J. H. Gillard. Identifying inflamed carotid plaques using in vivo USPIO-enhanced MR imaging to label plaque macrophages. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 26(7):1601–1606, 2006.

[204] R. A. Trivedi, J.-M. U-King-Im, M. J. Graves, J. J. Cross, J. Horsley, M. J. Goddard, J. N. Skepper, G. Quartey, E. Warburton, I. Joubert, L. Wang, P. J. Kirkpatrick, J. Brown, and J. H. Gillard. In vivo detection of macrophages in human carotid atheroma: Temporal dependence of ultrasmall superparamagnetic particles of iron oxide-enhanced MRI. *Stroke*, 35(7):1631–1635, 2004.

[205] M. T. B. Truijman, M. E. Kooi, A. C. van Dijk, A. A. J. de Rotte, A. G. van der Kolk, M. I. Liem, F. H. B. M. Schreuder, B. E., W. H. Mess, R. J. van Oostenbrugge, P. J. Koudstaal, L. J. Kappelle, P. J. Nederkoorn, A. J. Nederveen, J. Hendrikse, A. F. W. van der Steen, M. J. A. P. Daemen, and A. van der Lugt. PARISK (Plaque At RISK): Prospective multicenter study to improve diagnosis of high risk carotid plaques. *International Journal of Stroke*, 2013. epub ahead of print.

[206] N. Tustison, B. Avants, P. Cook, Y. Zheng, A. Egan, P. Yushkevich, and J. Gee. N4ITK: Improved N3 bias correction. *Medical Imaging, IEEE Transactions on*, 29(6):1310–1320, 2010.

[207] J. M. U-King-Im, A. J. Fox, R. I. Aviv, P. Howard, R. Yeung, A. R. Moody, and S. P. Symons. Characterization of carotid plaque hemorrhage: A CT angiography and MR intraplaque hemorrhage study. *Stroke*, 41(8):1623–1629, 2010.

[208] E. Ukwatta, J. Yuan, D. Buchanan, B. Chiu, J. Awad, W. Qiu, G. Parraga, and A. Fenster. Three-dimensional segmentation of three-dimensional ultrasound carotid atherosclerosis using sparse field level sets. *Medical Physics*, 40(5):052903, 2013.

[209] E. Ukwatta, J. Yuan, M. Rajchl, W. Qiu, D. Tessier, and A. Fenster. 3-D carotid multi-region MRI segmentation by globally optimal evolution of coupled surfaces. *IEEE Transactions on Medical Imaging*, 32(4):770–785, 2013.

[210] Q. J. A. van den Bouwhuijsen, M. W. Vernooij, A. Hofman, G. P. Krestin, A. van der Lugt, and J. C. M. Witteman. Determinants of magnetic resonance imaging detected carotid plaque components: the Rotterdam study. *European Heart Journal*, 33(2):221–229, 2012.

[211] F. van der Lijn, M. de Bruijne, S. Klein, T. den Heijer, Y. Y. Hoogendam, A. van der Lugt, M. M. B. Breteler, and W. J. Niessen. Automated brain structure segmentation based on atlas registration and appearance models. *IEEE Transactions on Medical Imaging*, 31(2):276–286, 2012.

[212] R. A. van Dijk, R. Virmani, J. H. von der Thüsen, A. F. Schaapherder, and J. H. N. Lindeman. The natural history of aortic atherosclerosis: A systematic histopathological evaluation of the peri-renal region. *Atherosclerosis*, 210(1):100–106, 2010.

[213] A. van Engelen, M. de Bruijne, S. Klein, H. J. M. Verhagen, H. C. Groen, J. J. Wentzel, A. van der Lugt, and W. Niessen. Plaque characterization in ex vivo MRI evaluated by dense 3D correspondence with histology. In *Proc. of SPIE Vol. 7963, 7963-80*, 2011.

[214] A. van Engelen, W. J. Niessen, S. Klein, H. C. Groen, H. J. M. Verhagen, J. J. Wentzel, A. van der Lugt, and M. de Bruijne. Supervised in-vivo plaque characterization incorporating class label uncertainty. In *9th IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 246 –249, may 2012.

[215] M. J. van Gils, K. Hameeteman, M. Straten, W. J. Niessen, and A. Lugt. Quantitative CT imaging of carotid arteries. In L. Saba, J. M. Sanches, L. M. Pedro, and J. S. Suri, editors, *Multi-Modality Atherosclerosis Imaging and Diagnosis*, pages 129–144. Springer New York, 2014.

[216] A. van Opbroek, M. A. Ikram, M. W. Vernooij, and M. de Bruijne. Supervised image segmentation across scanner protocols: A transfer learning approach. In *MICCAI 2012 Workshop on Machine Learning in Medical Imaging*, 2012.

[217] A. van Opbroek, M. A. Ikram, M. W. Vernooij, and M. de Bruijne. A transfer-learning approach to image segmentation across scanners by maximizing distribution similarity. In *MICCAI 2013 Workshop on Machine Learning in Medical Imaging*, 2013.

[218] R. van 't Klooster, P. J. H. de Koning, R. A. Dehnavi, J. T. Tamsma, A. de Roos, J. H. C. Reiber, and R. J. van der Geest. Automatic lumen and outer wall segmentation of the carotid artery using deformable three-dimensional models in MR angiography and vessel wall images. *Journal of Magnetic Resonance Imaging*, 35(1):156–165, 2012.

[219] R. van 't Klooster, O. Naggara, R. Marsico, J. H. C. Reiber, J.-F. Meder, R. J. van der Geest, E. Touzé, and C. Oppenheim. Automated versus manual in vivo segmentation of carotid plaque MRI. *American Journal of Neuroradiology*, 33(8):1621–1627, 2012.

[220] R. van 't Klooster, M. Staring, S. Klein, R. M. Kwee, M. E. Kooi, J. H. C. Reiber, B. P. F. Lelieveldt, and R. J. van der Geest. Automated registration of multispectral MR vessel wall images of the carotid artery. *Medical Physics*, 40(12):121904, 2013.

[221] M. O. Versteylen, B. L. Kietselaer, P. C. Dagnelie, I. A. Joosen, A. Dedic, R. H. Raaijmakers, J. E. Wildberger, K. Nieman, H. J. Crijns, W. J. Niessen, M. J. Daemen, and L. Hofstra. Additive value of semiautomated quantification of coronary artery disease using cardiac computed tomographic angiography to predict future acute coronary syndrome. *Journal of the American College of Cardiology*, 61(22):2296–2305, 2013.

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

[223] R. Virmani, F. D. Kolodgie, A. P. Burke, A. V. Finn, H. K. Gold, T. N. Tulenko, S. P. Wrenn, and J. Narula. Atherosclerotic plaque progression and vulnerability to rupture: Angiogenesis as a source of intraplaque hemorrhage. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 25(10):2054–2061, 2005.

[224] S. Voros, S. Rinehart, Z. Qian, G. Vazquez, H. Anderson, L. Murrieta, C. Wilmer, H. Carlson, K. Taylor, W. Ballard, D. Karmpaliotis, A. Kalynych, and C. B. III. Prospective validation of standardized, 3-dimensional, quantitative coronary computed tomographic plaque measurements using radiofrequency backscatter intravascular ultrasound as reference standard in intermediate coronary arterial lesions: Results from the ATLANTA (assessment of tissue characteristics, lesion morphology, and hemodynamics by angiography with fractional flow reserve, intravascular ultrasound and virtual histology, and noninvasive computed tomography in atherosclerotic plaques) I study. *JACC: Cardiovascular Interventions*, 4(2):198–208, 2011.

[225] P. C. Vos, T. Hambrock, C. A. H. van de Kaa, J. J. Fütterer, J. O. Barentsz, and H. J. Huisman. Computerized analysis of prostate lesions in the peripheral zone using dynamic contrast enhanced MRI. *Medical Physics*, 35(3):888–899, 2008.

[226] D. Vukadinovic, S. Rozie, M. Gils, T. Walsum, R. Manniesing, A. Lugt, and W. J. Niessen. Automated versus manual segmentation of atherosclerotic carotid plaque volume and components in CTA: associations with cardiovascular risk factors. *The International Journal of Cardiovascular Imaging*, 28(4):877–887, 2012.

[227] D. Vukadinovic, T. van Walsum, R. Manniesing, S. Rozie, K. Hameeteman, T. T. de Weert, A. van der Lugt, and W. Niessen. Segmentation of the outer vessel wall of the common carotid artery in CTA. *IEEE Transactions on Medical Imaging*, 29(1):65–76, 2010.

[228] T. J. Wang. Assessing the role of circulating, genetic, and imaging biomarkers in cardiovascular risk prediction. *Circulation*, 123(5):551–565, 2011.

[229] T. Wannarong, G. Parraga, D. Buchanan, A. Fenster, A. A. House, D. G. Hackam, and J. D. Spence. Progression of carotid plaque volume predicts cardiovascular events. *Stroke*, 44(7):1859–1865, 2013.

[230] A. D. Ward, C. Crukley, C. A. McKenzie, J. Montreuil, E. Gibson, C. Romagnoli, J. A. Gomez, M. Moussa, J. Chin, G. Bauman, and A. Fenster. Prostate: Registration of digital histopathologic images to in vivo MR images acquired by using endorectal receive coil. *Radiology*, 263(3):856–864, 2012.

[231] B. A. Wasserman, W. I. Smith, H. H. Trout, R. O. Cannon, R. S. Balaban, and A. E. Arai. Carotid artery atherosclerosis: In vivo morphologic characterization with gadolinium-enhanced double-oblique MR imaging - initial results. *Radiology*, 223(2):566–573, 2002.

[232] X. Wei. *Gray Level Run Length Matrix Toolbox v1.0.* Beijing Aeronautical Technology Research Center, 2007.

[233] J. Weickert. Multiscale texture enhancement. In V. Hlavác and R. Šára, editors, *Computer Analysis of Images and Patterns*, volume 970 of *Lecture Notes in Computer Science*, pages 230–237. Springer Berlin / Heidelberg, 1995.

[234] J. S. Weszka, C. R. Dyer, and A. Rosenfeld. A comparative study of texture measures for terrain classification. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-6(4):269 –285, 1976.

[235] W. H. O. (WHO). *The global burden of disease: 2004 update.* World Health Organization, Geneva, 2008.

[236] J. E. Wilhjelm, M.-L. M. Gronholdt, B. Wiebe, S. K. Jespersen, L. K. Hansen, and H. Sillesen. Quantitative analysis of ultrasound B-mode images of carotid atherosclerotic plaque: correlation with visual classification and histological examination. *IEEE Transactions on Medical Imaging*, 17(6):910–922, 1998.

[237] M. Wintermark, S. S. Jawadi, J. H. Rapp, T. Tihan, E. Tong, D. V. Glidden, S. Abedin, S. Schaeffer, G. Acevedo-Bolton, B. Boudignon, B. Orwoll, X. Pan, and D. Saloner. High-resolution CT imaging of carotid artery atherosclerotic plaques. *American Journal of Neuroradiology*, 29(5):875–882, 2008.

[238] P. Wu and T. Dietterich. Improving SVM accuracy by training on auxiliary data sources. In *Proceedings of the twenty-first international conference on Machine learning*, page 110. ACM, 2004.

[239] K. Yamada, Y. Song, D. Hippe, J. Sun, L. Dong, D. Xu, M. Ferguson, B. Chu, T. Hatsukami, M. Chen, C. Zhou, and C. Yuan. Quantitative evaluation of high intensity signal on MIP images of carotid atherosclerotic plaques from routine TOF-MRA reveals elevated volumes of intraplaque hemorrhage and lipid rich necrotic core. *Journal of Cardiovascular Magnetic Resonance*, 14(1):81, 2012.

[240] Q. Yang, J. Liu, S. R. S. Barnes, Z. Wu, K. Li, J. Neelavalli, J. Hu, and E. M. Haacke. Imaging the vessel wall in major peripheral arteries using susceptibility-weighted imaging. *Journal of Magnetic Resonance Imaging*, 30(2):357–365, 2009.

[241] X. Yang, J. Jin, M. Xu, H. Wu, W. He, M. Yuchi, and M. Ding. Ultrasound common carotid artery segmentation based on active shape model. *Computational and Mathematical methods in medicine*, 2013:345968, 2013.

[242] S. Yoshimura, K. Yamada, M. Kawasaki, T. Asano, M. Kanematsu, M. Miyai, Y. Enomoto, Y. Egashira, and T. Iwama. Selection of carotid artery stenting or endarterectomy based on magnetic resonance plaque imaging reduced periprocedural adverse events. *Journal of Stroke and Cerebrovascular Diseases*, 22(7):1082–1087, 2012.

[243] C. Yuan, L. M. Mitsumori, M. S. Ferguson, N. L. Polissar, D. Echelard, G. Ortiz, R. Small, J. W. Davies, W. S. Kerwin, and T. S. Hatsukami. In vivo accuracy of multispectral magnetic resonance imaging for identifying lipid-rich necrotic cores and intraplaque hemorrhage in advanced human carotid plaques. *Circulation*, 104(17):2051–2056, 2001.

[244] C. Yuan, S.-X. Zhang, N. L. Polissar, D. Echelard, G. Ortiz, J. W. Davis, E. Ellington, M. S. Ferguson, and T. S. Hatsukami. Identification of fibrous cap rupture with magnetic resonance imaging is highly associated with recent transient ischemic attack or stroke. *Circulation*, 105(2):181–185, 2002.

[245] H. Zhao, X. Zhao, X. Liu, Y. Cao, D. S. Hippe, J. Sun, F. Li, J. Xu, and C. Yuan. Association of carotid atherosclerotic plaque features with acute ischemic stroke: A magnetic resonance imaging study. *European Journal of Radiology*, 82(9):e465–e470, 2013.

[246] X. Zhao, H. R. Underhill, C. Yuan, M. Oikawa, L. Dong, H. Ota, T. S. Hatsukami, Q. Wang, L. Ma, and J. Cai. Minimization of MR contrast weightings for the comprehensive evaluation of carotid atherosclerotic disease. *Investigative Radiology*, 45(1):36–41, 2010.

[247] Y. Zhuge and J. Udupa. Intensity standardization simplifies brain MR image segmentation. *Computer Vision and Image Understanding*, 113(10):1095–1103, 2009.

Samenvatting

Nederlandse Samenvatting

Inleiding

Hart- en vaatziekten komen veel voor en behoren tot de belangrijkste doodsoorzaken wereldwijd. Aderverkalking, ook wel atherosclerose genoemd, is één van de belangrijkste onderliggende processen die tot hart- en herseninfarcten leiden. De ontwikkeling van atherosclerose is een langzaam en lang onopgemerkt proces: een infarct is vaak pas het eerste symptoom. Ook is het bij patiënten waarvan bekend is dat ze atherosclerose hebben nog erg moeilijk om het risico op een infarct in te schatten. Daarom is het belangrijk om methodes te ontwikkelen waarmee een betere risicoschatting gemaakt kan worden bij patiënten met onderliggende risicofactoren of bekende aanwezigheid van atherosclerose. In dit proefschrift heb ik mij gericht op de halsslagader, en specifiek op de geautomatiseerde analyse van scans om de aangetaste vaatwand te bestuderen. Na een uitgebreidere beschrijving van de achtergrond licht ik hier de studies en resultaten uit dit proefschrift toe.

Atherosclerose en de halsslagader

De halsslagaders bevinden zich aan beide zijden van de hals en voorzien het hoofd van bloed (zie figuur S1a). Bij de bifurcatie splitsen de aders in de externe halsslagader die naar het gezicht loopt, en de interne halsslagader, die de hersenen van bloed voorziet. Precies deze bifurcatie is één van de locaties in het lichaam waar een grotere kans is op de vorming van atherosclerose.

Atherosclerose wordt meestal pas op latere leeftijd ontdekt, maar de ontwikkeling begint vaak al veel eerder. Er wordt aangenomen dat dit proces start als ontstekingsreactie van de cellen in de vaatwand op cholesterolmoleculen in het bloed. In dit proces hopen zich steeds meer cholesterol en afweercellen op in de vaatwand, die eerst naar buiten toe verdikt, maar in een later stadium ook naar binnen waardoor de bloeddoorstroming belemmerd wordt (zie figuur S1b). Deze door atherosclerose aangetaste verdikte vaatwand wordt de *plaque* genoemd. Op een gegeven moment kunnen ook andere effecten optreden zoals het lekken van bloed de plaque in (intraplaque bloedingen) of verkalking.

Algemene risicofactoren zijn onder andere leeftijd, bloeddruk, roken, cholesterolniveau en gewicht. De preciese locatie waar atherosclerosevorming plaatsvindt hangt af van de geometrie van het bloedvat. De halsslagader, samen met de kransslagaders van het hart, en de vaten naar de benen zijn de locaties met het grootste risico op plaquevorming. Omdat atherosclerose in grote mate afhankelijk is van risicofactoren die in het hele lichaam gelijk zijn, zie je vaak atherosclerose in meerdere vaten. Hierdoor kan het bestuderen van de halsslagader ook een indicatie geven van de ziekteprocessen elders in het lichaam.

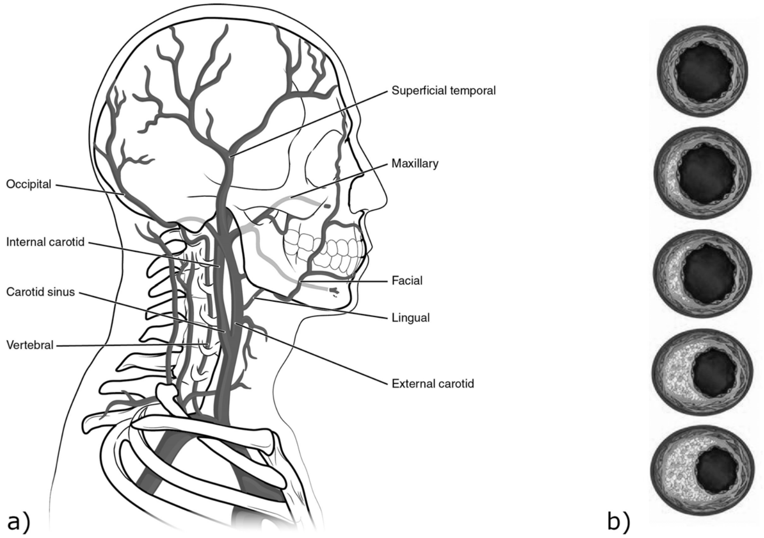


Figure S1: a) De vaten in het hoofd, waaronder de halsslagader met bifucatie. b) De ontwikkeling van atherosclerose van boven naar beneden: eerst verdikking van het vat naar buiten, met in een later stadium ook vernauwing van het vat naar binnen.

Risico van atherosclerose

In het geval van de halsslagader is het grootste risico niet dat een vat dichtslibt. In principe kan het andere vat het hele hoofd van bloed voorzien. Een groter risico is dat weefsel uit de plaque losraakt, via de bloedstroom meegevoerd wordt en verderop kleinere vaatjes kan verstoppen waardoor het achterliggende hersenweefsel geen bloed meer krijgt. Wanneer dit kortstondig gebeurt zonder blijvende gevolgen noemen we dit een TIA, wanneer het zuurstoftekort langer duurt en de gevolgen blijvend zijn is dit een herseninfarct. Het losraken of scheuren van de plaque noemen we plaque-ruptuur.

Voor een lange tijd is gedacht dat het risico op plaque-ruptuur afhangt van de mate van vernauwing van het bloedvat en/of verdikking van de vaatwand. Het is echter steeds duidelijker geworden dat andere kenmerken van de plaque belangrijkere factoren zijn. Door het vergelijken van weefsel uit de vaatwand van patiënten die wel of geen infarct gehad hebben is gevonden dat plaques een groter risico vormen wanneer er een grotere vetkern van cholesterol is, een dunnere laag weefsel (de fibreuze kap) deze kern van de bloedstroom scheidt, meer ontstekingscellen en intraplaque bloedingen aanwezig zijn en er juist minder kalkvorming is. Ook in studies waarbij patiënten langere tijd gevolgd zijn blijken deze kernmerken een belangrijke voorspellende waarde te hebben. Deze kernmerken maken de plaque instabieler, waardoor de kans groter is dat de krachten op bijvoorbeeld de overliggende fibreuze kap zo groot worden dat deze scheurt en het weefsel uit de plaque in de bloedstroom kan geraken.

Behandeling

In een vroeg stadium worden vaak middelen als bloedverdunners, en bloeddruk- en cholesterolverlagende middelen voorgeschreven. In een later stadium is ook een operatie mogelijk. Op dit moment worden patiënten waarvan bekend is dat ze een vaatvernauwing van meer dan 70% hebben geopereerd. Hierbij wordt dan meestal de plaque uit het vat weggesneden. Verschillende bronnen geven aan dat het de moeite waard is om andere parameters te bestuderen om tot nieuwe richtlijnen te komen die een beter onderscheid kunnen maken tussen patiënten die een hoog of laag risico lopen. De compositie van de plaque is hierbij een belangrijke parameter.

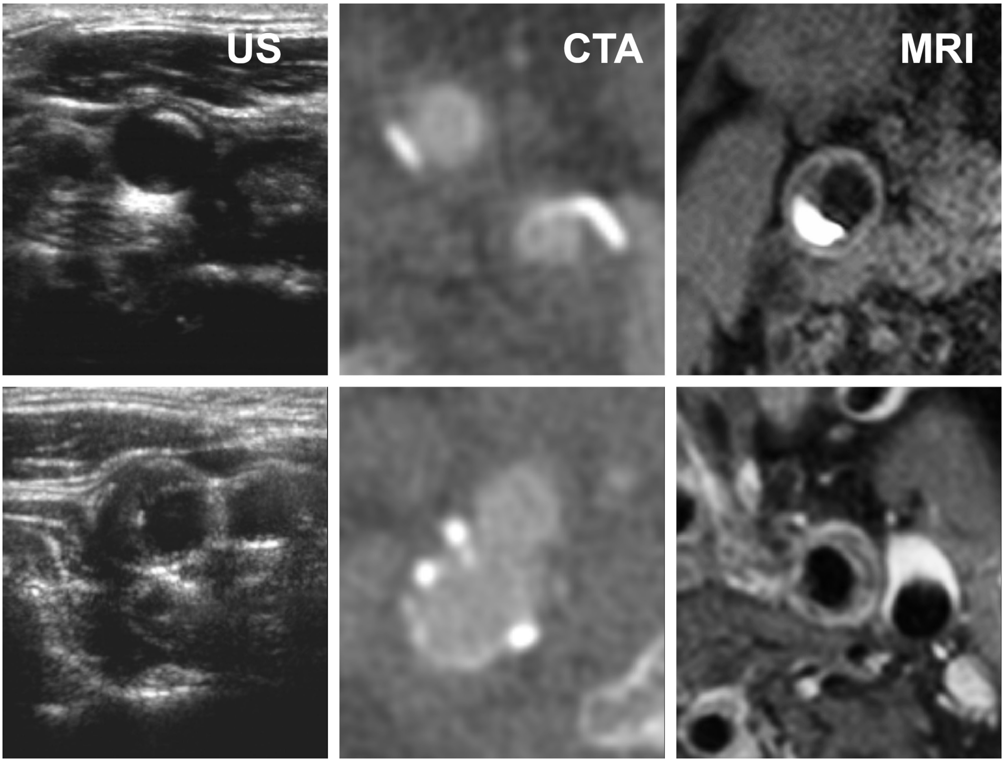


Figure S2: Voorbeelden van scans van halsslagaders met aderverkalking: Echo (US), CT-angiografie (CTA) en MRI. In het bovenste echobeeld heeft de plaque een hoge intensiteit, wat betekent dat het waarschijnlijk uit fibreus weefsel bestaat. De plaque in de onderste figuur heeft een lage intensiteit wat wijst op de aanwezigheid van een grote vetkern. In de CTA is kalk duidelijk te zien als witte gebieden. De vaatholte is ook duidelijk te herkennen, maar het onderscheid tussen de wand en het omliggende weefsel is lastiger te zien. De MRI in de bovenste rij heeft een gebied met intraplaque bloedingen met hoge intensiteit en een kalkregio met lage intensiteit vlak daarboven. De onderste rij laat een MRI na contrastinjectie zien waar het gebied midden in de verdikte wand donker is ten opzichte van de rand vanwege een vetkern.

Beeldvorming van de vaatwand

Om de samenstelling van de vaatwand te bepalen bij patiënten met atherosclerose wordt beeldvorming door middel van scans gebruikt. Er zijn verschillende beeldvormende technieken om dit te doen. De meestgebruikte niet-invasieve methodes zijn echo, CT en MRI. Deze drie technieken worden ook in dit proefschrift gebruikt. Voorbeelden van beelden van de halsslagader gemaakt met deze drie methoden zijn te zien in Figuur S2. Echo is van de drie het goedkoopste. Een onderzoeker kan het apparaat vasthouden en op de hals van de patiënt plaatsen. Verdikking van de vaatwand is duidelijk zichtbaar, en ook het verschil tussen een plaque met veel fibreus weefsel (normale vaatwand) en/of kalk en een plaque met meer vet en bloedingen is zichtbaar (zie de figuur). In tegenstelling tot traditionele 2D echobeelden hebben wij gebruik gemaakt van 3D reconstructies.

Voor CT en MRI wordt de patiënt in het apparaat geplaatst. Bij CT wordt een 3D beeld gevormd door het gebruik van radioactieve straling. De interactie van de straling met weefsel verschilt per weefsel, waardoor plaatjes gevormd kunnen worden. Vooral kalk is erg duidelijk zichtbaar op CT; de verschillen tussen fibreus en vetweefsel zijn lastiger te zien. MRI maakt gebruik van een sterk magneetveld. Doordat verschillende weefsels verschillende magnetische eigenschappen hebben worden deze weefsels verschillend in beeld gebracht. Voordelen van MRI zijn dat er geen schadelijke straling wordt gebruikt en dat ook zachte weefsels goed onderscheiden kunnen worden. Van zowel CT als MRI zijn in dit proefschrift ook beelden gebruikt die na injectie van contrastvloeistof gemaakt zijn. Bij CT-angiografie (CTA) zijn de vaatholtes waar het bloed doorheen stroomt dan duidelijker zichtbaar en is er iets beter een onderscheid tussen fibreus weefsel en vet te maken, en bij MRI helpt contrast bij het onderscheiden van de vetkern.

Geautomatiseerde beeldanalyse

Om maten van vaatwandcompositie te kunnen gebruiken in de praktijk is het naast het bepalen van aan- of afwezigheid van componenten, ook belangrijk deze te kwantificeren: hoe groot is de vetkern, hoeveel kalk is er, etc. Deze maten kun je verkrijgen door een arts in een 3D-beeld van de vaatwand in elke 2D plak contouren te laten trekken om de verschillende componenten. Dit kost echter heel veel tijd en het is daarom onmogelijk om dit in de praktijk op grote schaal toe te passen. Een andere mogelijkheid is om dit werk (gedeeltelijk) door een computer te laten doen. Het ontwikkelen van computeralgoritmes die de verschillende plaque-componenten kunnen herkennen en kwantificeren was het doel van dit proefschrift.

Er zijn verschillende manieren om dit aan te pakken, maar in dit proefschrift is vooral gebruik gemaakt van voxel-classificatie. Een voxel is een pixel in 3D, en dus een klein blokje in een 3D-beeld. Bij voxel-classificatie wordt van elk voxel in een beeld, of een bepaald gedeelte van een beeld, bepaald tot welke klasse de voxel hoort. Dit kan bijvoorbeeld zijn ‘kalk’, ‘vet’, ‘intraplaque bloeding’ of ‘fibreus weefsel’. Zo verkrijg je een segmentatie van je beeld waarin je de grootte van de verschillende componenten kunt bepalen door voxels te tellen.

Het classificeren wordt gedaan op basis van kenmerken die je kunt meten, zoals de intensiteit in een MRI-scan, of de afstand tot de vaatholte. Om een model te ontwikkelen is een grote set met voorbeelden nodig op basis waarvan de kenmerken van iedere klasse bepaald kunnen worden. Een voorbeeld van hoe dit gaat is te zien in figuur S3A. Elke stip in

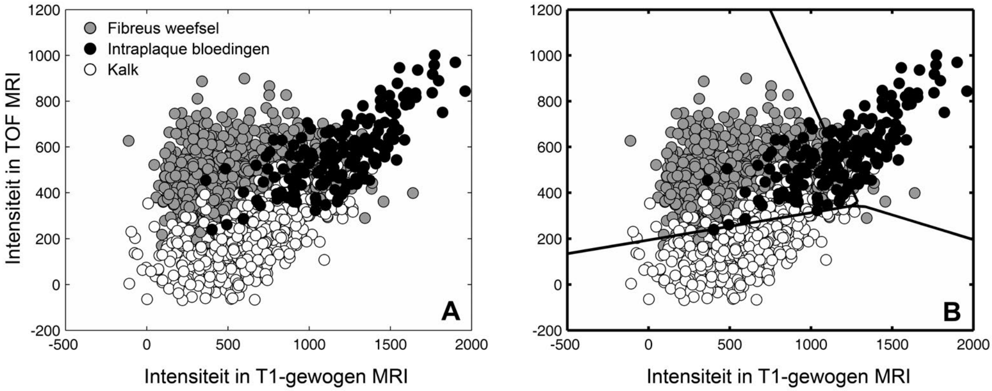


Figure S3: Voorbeeld van het maken van een classificatiemodel. A) Van elke voxel uit een trainingsset is de intensiteit in twee MRI-scans bepaald en het label van de juiste klasse. B) Hier zijn ook de optimale grenzen die gebruikt kunnen worden om voxels uit een nieuw beeld te classificeren als één van de klassen aangegeven.

deze figuur vertegenwoordigt een voxel uit de vaatwand. De positie van elk voxel in de figuur is bepaald aan de hand van de intensiteit in één scan, de T1-gewogen MRI, en een andere scan, de TOF MRI. De kleur van de stip hangt af van het label dat het voxel gekregen heeft. Dit kan bijvoorbeeld gedaan zijn door een arts die verschillende MRI-scans van dezelfde patiënt naast elkaar heeft gelegd en voor het bloedvat contouren heeft getrokken om de binnen- en buitenwand en de verschillende componenten in de wand. Je kunt in deze figuur zien dat in de T1-gewogen MRI geen duidelijk onderscheid is tussen kalk en fibreus weefsel, maar dat intraplaque bloedingen gemiddeld een hogere intensiteit hebben. In de TOF-scan daarentegen zie je wel een duidelijk verschil tussen kalk en fibreus weefsel. Intraplaque bloedingen hebben hier een gelijke of hogere intensiteit dan het fibreuze weefsel. Het is dus duidelijk dat er meerdere scans nodig zijn om de verschillende componenten goed te kunnen onderscheiden. In de praktijk worden veel meer dan twee kenmerken gebruikt en heb je een hoog-dimensionale ruimte waar alle voxels in geplaatst worden. Zo'n set voorbeelden met labels wordt de trainingsset genoemd.

De volgende stap is om op basis van deze trainingsset te bepalen waar de optimale grens tussen de verschillende klassen loopt, zoals in figuur S3B. Er zijn veel modellen waarmee deze grens bepaald kan worden, en in ieder afzonderlijk geval zal een ander model het beste zijn, afhankelijk van de onderliggende data. Voor de studies in dit proefschrift is meestal een model gemaakt op basis van 1) het gemiddelde per klasse voor elk kenmerk, 2) de spreiding van elke klasse in de verschillende richtingen, en 3) de vooraf te verwachten kans op elke klasse zonder dat je specifieke kenmerken van het vat meeneemt (in de meeste vaten is meer fibreus weefsel aanwezig dan kalk, vet, of bloedingen). Hierbij is als voorwaarde gesteld dat de grenzen recht moeten zijn, waardoor in het voorbeeld de grenzen lopen zoals te zien is in figuur S3B.

In scans van nieuwe patiënten kunnen nu van alle voxels alle kenmerken bepaald worden. Zo wordt er van ieder voxel bepaald aan welke kant van de grens deze voxels ligt en dus tot welke klasse het hoogstwaarschijnlijk behoort.

Je ziet in figuur S3B dat er geen perfecte scheiding is tussen de verschillende componenten. Dit kan verschillende redenen hebben. Zo is het belangrijk dat het vat precies op dezelfde plek ligt in de verschillende scans die gemaakt worden van een patiënt. Omdat de patiënt meestal wel wat beweegt tussen het maken van twee scans door moet hiervoor gecorrigeerd worden door sommige scans iets te verplaatsen zodat ze allemaal gelijk komen te liggen. Dit heet beeldregistratie. Daarnaast heeft MRI geen gestandaardiseerde waardes. In elke CTA-scan is de intensiteit van bloed of kalk gelijk, maar bij MRI verschilt dit van moment tot moment. Hiervoor zijn verschillende normalizatiemethodes beschikbaar. Ook is de intensiteit niet overal in het beeld gelijk voor hetzelfde weefsel, dus ook hier moet voor gecorrigeerd worden. Ook wanneer dit allemaal gedaan is blijft er vaak overlap tussen de klassen, onder andere door ruis en doordat de labels vaak niet perfect zijn. Hierdoor is het altijd belangrijk classificatiemethodes te testen op data die onafhankelijk is van de trainingsset. Als een model te specifiek op de trainingsset past, modelleert het kleine details zoals ruis te specifiek waardoor je fouten krijgt in nieuwe data.

Bijdragen uit dit proefschrift

Voor dit proefschrift zijn verschillende studies uitgevoerd. In hoofdstuk 2-4 hebben we gebruik gemaakt van een dataset van patiënten die een operatie hebben ondergaan waarbij de plaque uit de halsslagader verwijderd is. Voor de operatie zijn deze patiënten gescand met MRI en CTA. Na de operatie zijn *ex vivo* scans met heel hoge resolutie gemaakt van de plaque, die vervolgens in hele dunne plakken gesneden is (histologie). Van deze histologiebeelden is onder de microscoop heel nauwkeurig te bepalen welke weefsels aanwezig zijn, en dit is vervolgens gebruikt om de labels te bepalen om segmentatiemethoden te ontwikkelen.

In hoofdstuk 2 hebben we gegeken naar de *ex vivo* MRI-scans. Hier laten we zien dat het gebruiken van extra kenmerken naast MRI-intensiteit en de positie binnen de vaatwand tot betere segmentaties leidt. Deze kenmerken zijn wiskundige afgeleides in het beeld, waarmee je kijkt naar veranderingen in intensiteit. Ook laten we zien dat het voordelen kan hebben om de registratie met histologie op een uitgebreide 3D-manier te doen in plaats van het in 2D aan elkaar linken van plakken uit de beelden.

In hoofdstuk 3 hebben we naar de MRI- en CTA-scans gekeken die gemaakt zijn voor de operatie. Het probleem hier was dat het vanwege vervormingen van de plaque tijdens de operatie en bij het snijden van het weefsel vaak onmogelijk is om de scans precies te registreren met histologie. Daarom hebben we in deze studie verschillende manieren bedacht om tijdens het trainen van het classificatiemodel rekening te houden met fouten in de registratie. Hierbij bepaalden we voor de voxels in de hoogdimensionale ruimte naast het label ook hoe betrouwbaar dit label is aan de hand van hoe goed de registratie gelukt was, of gaven we het een label voor verschillende klassen met een bepaalde kans op de grens tussen structuren. Het bleek echter het beste te werken om de voorbeelden die het meest afwijken van het gemiddelde van de klasse niet te gebruiken. Ook hebben we in dit hoofdstuk laten zien dat MRI goed gebruikt kan worden om fibreus en vetweefsel te onderscheiden, maar niet om kalk te kwantificeren en dat het voor CTA juist andersom is. Wanneer je ze beiden gebruikt kun je al deze drie componenten goed kwantificeren.

In hoofdstuk 2 en 3 is het belangrijk om een goede segmentatie van de histologiebeelden te hebben, omdat dit gebruikt wordt om de labels in de trainingsset te bepalen. Omdat dit erg veel werk is hebben we in hoofdstuk 4 een classificatiemethode ontwikkeld om de histologiebeelden automatisch te segmenteren. We laten zien dat een volledig automatische methode redelijk nauwkeurige segmentaties geeft van de histologie, maar dat wanneer deze gebruikt worden om segmentatiemethoden zoals in hoofdstuk 2 en 3 te ontwikkelen de fouten groter zijn dan wanneer de handmatig ingetekende histologie gebruikt wordt. Wanneer van elke plaque één of twee histologiebeelden worden ingetekend om op te trainen zijn de histologiesegmentaties nauwkeuriger en krijg je ook goede segmentaties voor de *ex vivo* en *in vivo* beelden.

In hoofdstuk 5 gebruiken we scans van patiënten met atherosclerose die een vaatvernauwing kleiner dan 70% hebben, dus patiënten die niet geopereerd worden. Deze patiënten zijn gescand in verschillende ziekenhuizen met verschillende typen scanners. Meestal is het zo dat een methode ontwikkeld wordt voor een bepaalde dataset waarin de scanner en de instellingen van de scanner hetzelfde zijn. Wanneer je deze methode vervolgens wil gebruiken om data van een andere scanner te analyseren is de nauwkeurigheid over het algemeen veel lager. Wij hebben op twee manieren geprobeerd om een methode die ontwikkeld is op gelabelde data van het ene ziekenhuis met succes toe te passen op de data van het andere ziekenhuis. Zowel door een uitgebreidere normalizatie van de twee datasets toe te passen, als door slechts een paar (3-10) beeldplakken uit het andere ziekenhuis in te tekenen en toe te voegen aan de trainingsset uit het andere ziekenhuis, konden we goede segmentaties verkrijgen.

In hoofdstuk 6 hebben we echobeelden gebruikt. In plaats van segmentatie van plaque componenten, hebben we hier op een andere, indirecte, manier de samenstelling van de vaatwand proberen te bepalen. Van 298 patiënten met atherosclerose en een vaatvernauwing kleiner dan 70% zijn twee 3D-echoscans gemaakt met een tussenperiode van 1 jaar. In al deze scans zijn de plaques ingetekend en zijn per patiënt eigenschappen van de textuur van deze plaques bepaald. Voorbeelden zijn het gemiddelde contrast, homogeniteit en of de textuur grof of fijn is. Van deze patiënten was bekend wie een herseninfarct, TIA of hartinfarct heeft gekregen en wanneer. Met deze informatie hebben wij een model gemaakt dat verschillende textuurmaten zo combineert dat een redelijk goed onderscheid te maken is tussen patiënten met een hoog en laag risico. Het bleek dat het kijken naar verandering in textuur gecombineerd met de verandering in plaque volume de beste voorspelling geeft.

Toekomstperspectief

In dit proefschrift hebben we laten zien dat het mogelijk is om de verschillende componenten uit de atherosclerotische vaatwand te kwantificeren uit niet-invasieve beelden. Vervolgens zal, zoals in hoofdstuk 6, in studies waarbij patiënten langere tijd gevolgd worden, bepaald moeten worden of en welke van deze maten een goede voorspelling kunnen geven van het risico op plaque-ruptuur. Ook de patiënten uit de studie in hoofdstuk 5 worden na twee jaar opnieuw gescand en nog enkele jaren gevolgd om te zien wie symptomen als een herseninfarct ontwikkelt en wie niet. Op dit moment loopt de studie nog, maar de segmentatiemethode die in hoofdstuk 5 is ontwikkeld kan gebruikt worden om uiteindelijk de hele dataset te analyseren.

Vervolgens zal een klinische studie uitgevoerd moeten worden waarbij één groep patiënten op basis van de huidige richtlijnen behandeld wordt, en één groep patiënten op basis van nieuwe criteria. Dit zal dan uit moeten wijzen of met maten van plaque-samenstelling gerichter patiënten geselecteerd kunnen worden voor behandeling. Hopelijk kan het onderzoek uit dit proefschrift hier een bijdrage aan leveren.

Publications

Journal publications

- **A. van Engelen**, T. Wannarong, G. Parraga, W.J. Niessen, A. Fenster, J.D. Spence and M. de Bruijne, *Three-dimensional carotid ultrasound plaque texture predicts vascular events*, submitted
- **A. van Engelen**, A.C. van Dijk, M.T.B. Truijman, R. van 't Klooster, A. van Opbroek, A. van der Lugt, W.J. Niessen, M.E. Kooi and M. de Bruijne, *Multi-center MRI plaque-component segmentation using feature normalization and transfer learning*, Submitted
- **A. van Engelen**, W.J. Niessen, S. Klein, H.C. Groen, H.J.M. Verhagen, J.J. Wentzel, A. van der Lugt and M. de Bruijne, *Atherosclerotic plaque component segmentation in combined carotid MRI and CTA data incorporating class label uncertainty*, PLOS ONE, 2014
- **A. van Engelen**, W.J. Niessen, S. Klein, H.C. Groen, K. van Gaalen, H.J.M. Verhagen, J.J. Wentzel, A. van der Lugt and M. de Bruijne, *Automated segmentation of atherosclerotic histology based on pattern classification*, Journal of Pathology Informatics, 2013
- **A. van Engelen**, W.J. Niessen, S. Klein, H.C. Groen, H.J.M. Verhagen, J.J. Wentzel, A. van der Lugt and M. de Bruijne, *Multi-feature-based plaque characterization in ex vivo MRI trained by registration to 3D histology*, Physics in Medicine and Biology, 2012
- P.D.H.M. Verhaegen, E.M. Res, **A. van Engelen**, E. Middelkoop and P.P.M. van Zuijlen, *A reliable, non-invasive measurement tool for anisotropy in normal skin and scar tissue*, Skin Research and Technology, 2010

Conference papers

- A. Arias Lorza, J. Petersen, **A. van Engelen**, H. Tang, M. Selwaness, J.C.M. Witteman, A. van der Lugt, W.J. Niessen and M. de Bruijne, *Carotid artery wall segmentation by coupled surface graph cuts*, MICCAI Medical Computer Vision Workshop 2012, 2012
- **A. van Engelen**, W.J. Niessen, S. Klein, H.C. Groen, K. van Gaalen, H.J.M. Verhagen, J.J. Wentzel, A. van der Lugt and M. de Bruijne, *Automated Segmentation of Atherosclerotic Histology Based on Pattern Classification*, MICCAI workshop on Histopathology Image Analysis: Image computing in digital pathology, 2012
- **A. van Engelen**, W.J. Niessen, S. Klein, H.C. Groen, H.J.M. Verhagen, J.J. Wentzel, A. van der Lugt and M. de Bruijne, *Supervised in-vivo plaque characterization incorporating class label uncertainty*, 9th IEEE International Symposium on Biomedical Imaging 2012, 2012

- **A. van Engelen**, M. de Bruijne, S. Klein, H.J.M. Verhagen, H.C. Groen, J.J. Wentzel, A. van der Lugt and W.J. Niessen, *Plaque characterization in ex vivo MRI evaluated by dense 3D correspondence with histology*, Proc. of SPIE Vol. 7963, 7963-80, 2011

Conference abstracts

- **A. van Engelen**, T. Wannarong, G. Parraga, W.J. Niessen, A. Fenster, J.D. Spence and M. de Bruijne, *Atherosclerotic plaque texture from 3D carotid ultrasound improves prediction of vascular events*, European Congress of Radiology, 2014
- **A. van Engelen**, A.C. van Dijk, M.T.B. Truijman, R. van 't Klooster, A. van der Lugt, W.J. Niessen, M.E. Kooi and M. de Bruijne, *Automatic segmentation of atherosclerotic plaque components in different carotid MRI protocols*, European Congress of Radiology, 2014
- **A. van Engelen**, A.C. van Dijk, M.T.B. Truijman, R. van 't Klooster, A. van der Lugt, W.J. Niessen, M.E. Kooi and M. de Bruijne, *Automatic segmentation of atherosclerotic plaque components in different carotid MRI protocols*, NVPHBV Fall meeting, 2013
- **A. van Engelen**, M. de Bruijne, S. Klein, H.C. Groen, J.J. Wentzel, A. van der Lugt and W.J. Niessen, *Plaque characterization in ex vivo MRI evaluated by dense 3D correspondence with histology*, 3^{rd} Dutch conference on Biomedical Engineering, 2011

Dankwoord

Acknowledgements

Een proefschrift schrijven doe je, gelukkig, niet alleen. Ik wil dan ook iedereen bedanken voor de inspiratie, motivatie, kennis, ideeën, afleiding en steun die ik gekregen heb.

Mijn promotor, beste Wiro, ik wil je bedanken voor zowel je warme en inspirerende persoonlijkheid, als voor je kennis van en inzicht in de wetenschap en het wereldje daar omheen. Ook al werden onze inhoudelijke besprekingen in de loop der tijd minder, telkens wist je direct de juiste punten te noemen om weer een stap verder te komen. Daarnaast wil ik bij deze de bewondering uitspreken voor de prachtige volzinnen die ik van jou terugkrijg nadat ik mijn stukken opstuur, precies op die plaatsen waar het mij niet wilde lukken.

Mijn copromotor, beste Marleen, bijna alles wat ik de afgelopen jaren heb geleerd over patroonherkenning en het doen van wetenschappelijk onderzoek in het algemeen heb ik van jou geleerd. Bedankt voor je geduld, inzichten, ideeën, en bedankt dat je mij ook de vrijheid gaf zelf te bedenken wat ik wilde doen. Je hebt me klaargestoomd om nu als zelfstandig onderzoeker de wijde wereld in te trekken, maar ik hoop dat we elkaar nog regelmatig tegen zullen komen.

I would like to thank the members of the reading committee of my thesis, prof. Ton van der Steen, prof. Bram van Ginneken and prof. Aaron Fenster for the time they took for reading my thesis. Dear Aaron, I would also like to thank you for the opportunity to perform part of my project in your research group at the Robarts Research Institute. I had a great time due to the opportunities you created and and the freedom you gave me. I am looking forward to meeting you again in Rotterdam for my defense.

Ik kon mijn onderzoek beginnen met een prachtige dataset die klaarlag voor gebruik. Harald, Kim, Jolanda, Stefan, bedankt hiervoor! Bedankt voor de betrokkenheid bij mijn project en de input die jullie zijn blijven leveren. Stefan, bedankt dat je zelfs toen ik voor de zoveelste keer langskwam om te vragen hoe de registratiepipeline ook alweer precies in elkaar zat het weer uitlegde!

Aad, vanaf het begin ben je vanuit de kliniek betrokken geweest bij mijn project. Bedankt voor je interesse en positieve instelling waardoor je me altijd het gevoel gaf dat de dingen waar ik aan werkte ook gewild zijn in de praktijk. Door je betrokkenheid en specifieke vragen hoop ik daar goed op in ingespeeld te kunnen hebben. Eline, voor mijn laatste project heb ik ook intensief met jou samengewerkt. Bedankt voor alle input die je gegeven hebt. Aad en Eline, ik vind het erg leuk dat jullie in mijn grote promotiecomissie zitten.

Doordat ik op een groot project werkte heb ik de kans gehad met veel mensen samen te werken. Allereerst de arts-onderzoekers, Anouk, Martine, Zaid, Floris: de samenwerking met jullie in vooral de laatste periode heeft een enorm positieve invloed op mijn project gehad. Ik weet hoeveel werk jullie in de sets contouren hebben gestoken, en ik hoop dan ook dat ik jullie in ruil daarvoor een mooie tool terug heb kunnen geven. De klinische studie is voorlopig nog niet ten einde, maar ik hoop dat dankzij ons werk er uiteindelijk erg mooie dingen uit gaan komen! Ook Alexandra en Madieke wens ik alvast veel succes hiermee.

Anouk, specifiek met jou heb ik veel samengewerkt en dit is me altijd heel goed bevallen, bedankt en heel veel succes met de laatste periode van jouw werk!

De mooie afronding van de MRI-analyse in een werkende tool was nooit gelukt zonder de samenwerking met Ronald en Rob. Ronald, ik vind het erg leuk dat je al die jaren naar het EMC kwam voor R-VIP. Naast de gezelligheid heeft het zeker tot positieve resultaten geleid in meerdere projecten, maar vooral bij de uiteindelijke samenwerking met betrekking tot VesselMass. Rob, ook hartelijk bedankt voor de samenwerking, telkens wanneer ik ergens tegenaan liep was het bijzonder snel geregeld.

Harm, ik ken bijna niemand die zoveel positieve energie heeft als jij. Na elk (korte) overleg barstte ik dan ook weer van de zin om aan nieuwe projecten te gaan werken. Al is er nooit een echt samenwerkingsproject gekomen, het was zeker nuttig met jou, en Frank hierbij zeker ook, ideeën te bespreken. Mocht je nog een keer een idee hebben, laat het vooral weten!

As I already mentioned, my visit to London, Canada, has been a very valuable experience. It wouldn't have been half as good without the people there. David, thanks for guiding me around campus during the first days, for helping me with anything thereafter, and for the good company during lunch and coffee breaks. Mohammad, thanks for the many discussions, squash games and beers. Mohammad and Jonathan, thanks for the trip to point Pelee, and the amazing winter camping trip! Also to all the other people at Robarts, thank you for the great time I had. My roommates Sarah and Tina, thank you for your company, for letting me experience some real Canadian traditions, and for the trip to Toronto.

Dan, Ton, David and Grace, thank you for working with me on the texture analysis project. Thanks to Lucy for continuing the work in this field, and thanks to Eranga and Jing for discussing your segmentation algorithms with me.

I would also like to thank everyone who contributed to my projects in the many R-VIP and MBM meetings over the past years. These meetings were really among the highlights of my weeks. Moreover, thank you for letting me ask lots of questions about your work and for making me learn from it.

I wouldn't have had such a good time if I hadn't had such a great group of colleagues. It is you who make BIGR such a good place to work. My roommates Diego and Andrés: I thank you for your company, for letting me be involved in your work, and for your introduction to South-American politics, viewpoints and habits! Annegreet, bedankt voor je inhoudelijke én morele steun bij de afronding van mijn laatste hoofdstuk (H5). Esther, ik vind het erg leuk dat we na studiegenoten ook collega's zijn geworden. Beiden bedankt voor jullie gezelligheid, ik vind het erg leuk dat jullie mijn paranimfen willen zijn! Desiree, bedankt voor het regelen van alle adminstratieve zaken rondom de promotie, dat heeft mij een hoop tijd en zorgen gescheeld. All other current and previous colleagues I would like to thank for your company during work, at conferences, and for all social activities (reflected in the recent explosion of BIGR Facebook groups). If you haven't had one, remember me of giving you a personal acknowledgement!

Ik had het genoegen vrienden binnen het Erasmus te hebben die niets met mijn werk te maken hebben: Caroline en Maarten, bedankt voor de koffiepauzes waarin ik ongestoord mijn frustraties kwijt kon of helemaal niet aan werk kon denken.

Anette en Kimmy, of het nu in het sportcentrum of het café is, met jullie is het altijd gezellig. Anette, bedankt voor je vrolijkheid, en je promotietips en -trucs. Kimmy, het was goed de periode van proefschrift afronden en solliciteren met je te kunnen delen. Harriët, ik vond het erg leuk de verschillende cursussen en SPIE in Florida samen met jou te kunnen doen, veel succes met je eigen proefschrift! Hans en Hanneke, of we nu gaan afspreken in Londen, Aarhus, Amsterdam of Ommen, laten we het vooral vaak doen! Barbara, Bianca en Hylkje: de sauna-/knutseldagen zijn altijd heerlijk ontspannend, binnenkort een keer in Berlijn? Jeroen, Walter, Dominique (Lars!), Linda, bedankt voor de gezellige spelletjesdagen, etentjes etc. Rosalie, het is altijd erg gezellig samen. Bedankt voor alle mailtjes over een weer waarin we allebei de ander kunnen zeggen 'het komt wel goed' :-)

Elmer, Elma en Femke, bedankt voor jullie gezelschap, betrokkenheid en interesse.

Papa en mama, het voelt bij jullie nog altijd helemaal als 'thuis' komen. Ik weet dat jullie altijd achter me staan, bedankt hiervoor! Luuk en Jettie, ik heb het maar getroffen met jullie als broer en zus. Het was leuk toeval dat we de afgelopen jaren alledrie in Utrecht woonden. Letten jullie erop dat ik mijn stem een beetje spaar twee dagen voor de verdediging?

Sjoerd, jouw steun is misschien wel het allerbelangrijkste geweest de afgelopen jaren. Bedankt dat je er altijd bent wanneer ik dat nodig heb. Je maakt mijn leven een stuk leuker, en ik kijk ernaar uit binnenkort weer in hetzelfde land te wonen!

PhD portfolio

PhD portfolio

In-depth courses

Advanced Pattern Recognition (ASCI)	2010
Measuring Features (ASCI)	2011
Front-end Vision and Multiscale Image Analysis (ASCI)	2011
Knowledge-Driven Image Segmentation (ASCI)	2012
C++ course (BIGR)	2011
Cardiovascular Image Analysis (COEUR)	2013
Principles of Research in Medicine (NIHES)	2010
Introduction to data-analysis (NIHES)	2010
Regression analysis (NIHES)	2010
Survival analysis (NIHES)	2010
Biomedical English Writing and Communication (Erasmus MC)	2011
Presentation course (Department of Medical Informatics)	2011
IEEE EMBS International Summer School on Biomedical Imaging, Berder	2010
Biomedical Image Analysis Summer School: Modalities, Methodologies & Clinical Research,Parijs	2013

International conferences (attendance/presentation)

IEEE International Symposium on Biomedical Imaging, Rotterdam (attendance)	2010
SPIE Medical Imaging, Orlando (poster presentation)	2011
European Congress of Radiology, Vienna (attendance)	2012
IEEE International Symposium on Biomedical Imaging, Barcelona (poster presentation)	2012
Medical Image Computing and Computer-Assisted Intervention, Nice (MICCAI, oral presentation at Histopathology Image Analysis workshop)	2012
European Congress of Radiology, Vienna (2 oral presentations)	2014

Attended workshops, seminars and symposia

Bayesian Methods and Bias Analysis (Erasmus MC)	2010
Machine Learning in Biomedical Imaging (ISBI)	2010
Graph algorithmic techniques for Biomedical Image Segmentation (SPIE)	2011
Several COEUR seminars	2010-2013
Several COEUR PhD days	2011-2013
Dutch Association for Pattern Recognition and Image Processing meeting	2011,2013
Medical Imaging Symposium for PhD students	2010,2011,2013
Netherlands Forum for Biomedical Imaging meeting, Leiden	2011
Imaging applications in prostate cancer, London (Canada)	2012
IMDI MedTech West, Delft	2013
Erasmus MC PhD Day	2011,2013

Other presentations

COEUR Research seminar (2x)	2010-2011
COEUR PhD Day	2011
Medical Image Symposium for PhD students	2013
Dutch Association for Pattern Recognition and Image Processing, fall meeting	2013
Internal BIGR Vascular Image Processing meetings	2010-2014
Internal BIGR Modelbased Image Analysis meetings	2010-2014
Medical Informatics Research Lunch	2010
BIGR seminar (3x)	2011-2013
Several Semi-annual Parisk meetings (oral/poster)	2010-2013
Several CTMM Annual Meetings (poster)	2010-2013

Research-related activities

Teaching Introduction to Image Processing to medical students	2011-2013
Organizing annual COEUR PhD Day	2012-2013
Participant "De Jonge Akademie on Wheels"(2x)	2011-2012
Reviewing for European Medical Physics, IEEE Transactions on Medical Imaging, Computational and Mathematical Methods in Medicine, and PLOS ONE	2012-2013

Curriculum V.

About the author

Arna van Engelen was born in Heesch, the Netherlands, in 1986. She obtained her vwo diploma from the Maaslandcollege in Oss in 2004. That year she started the university program *Medical Natural Sciences* at the VU university in Amsterdam. She obtained her Bachelor's degree in 2006 with a graduation project on the mathematical description of electrical tissue conductance for impedance measurements. She followed a Master's program in Medical Natural Sciences as well, with a specialization in Medical Physics. During this Master she performed a research project at the Dutch Association of Burn Centers on histological image analysis and measurements of collagen orientation in scar tissue. Her graduation project was done at the department of pulmonary diseases of the VU university medical center. Here she studied the contraction pattern of the heart of patients with pulmonary hypertension using MRI myocardial tagging. In 2010 she started her PhD project at the Biomedical Imaging Group Rotterdam, at Erasmus MC, of which the work is described in this thesis.

Arna van Engelen werd geboren in Heesch in 1986. Na het behalen van het vwo-diploma aan het Maaslandcollege te Oss in 2004, begon ze met de bachelorstudie Medische Natuurwetenschappen aan de Vrije Universiteit in Amsterdam. Haar bachelordiploma behaalde ze in 2006 met een afstudeerproject over de wiskundige beschrijving van weefselgeleiding. Hierna volgde ze een masteropleiding van dezelfde studie, met de specialisatie Medische Fysica. In deze periode deed zij een onderzoeksproject bij de Vereniging Samenwerkende Brandwondencentra Nederland naar collageenoriëntatie in littekenweefsel, met behulp van histologische beeldanalyse en weefselmetingen. Haar afstuderen deed zij op de afdeling longziekten van het VU Medisch Centrum, waar zij het contractiepatroon van het rechterventrikel bestudeerde bij patiënten met pulmonale hypertensie door de analyse van MRI-tagging beelden. In 2010 is ze begonnen aan haar promotieonderzoek bij de Biomedical Imaging Group Rotterdam, aan het Erasmus Medisch Centrum, waarvan het werk beschreven is in dit proefschrift.

ISBN: 978-90-6464-774-1