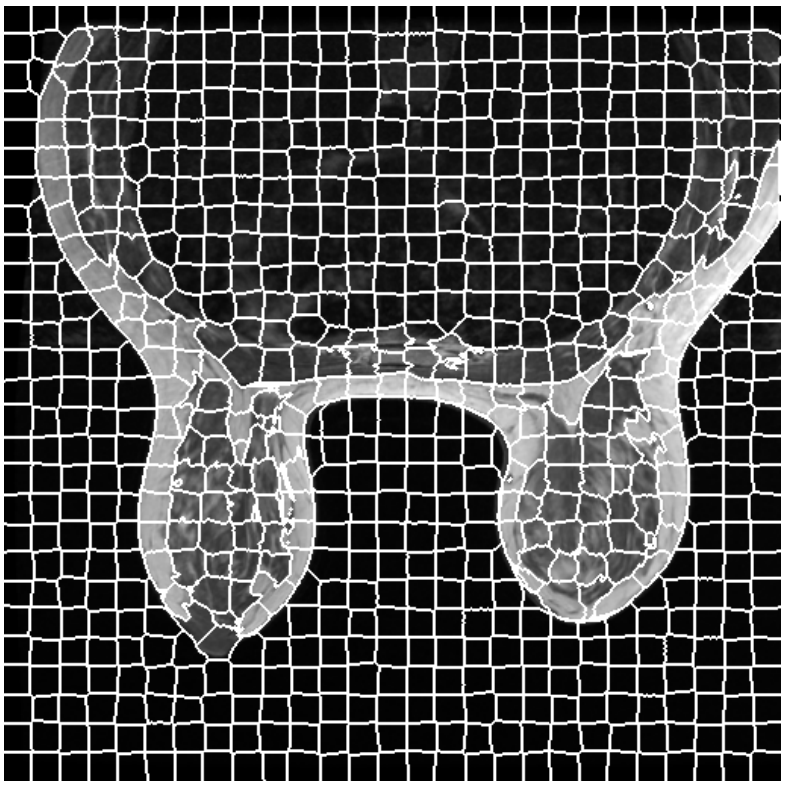
CHALMERS



Supervoxel-based algorithms for use in breast MRI CAD systems

Master of Science Thesis in Biomedical Engineering

MAGNUS ZIEGLER

Department of Signals and Systems
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden, 2014
Thesis EX045/2014

Supervoxel-based algorithms for use in breast MRI CAD systems

Magnus C. Ziegler

Technical report no. EX045/2014

Department of Signals and Systems

Chalmers University of Technology

SE-412 96 Göteborg

Sweden

Telephone: +46 (0)31-772 1000

Cover:

An axial image of supervoxels created from a breast Magnetic Resonance Image volume

Supervoxel-based algorithms for use in breast MRI CAD systems

Magnus C. Ziegler

September 30, 2014

Acknowledgments

Engineering is a team effort, and this work has been no different. I would like to thank my supervisor, Andrew Mehnert, for valuable discussions and guidance, Darryl McClymont, for help developing algorithms, and to the team at Medtech West who provided a fun working environment with many great fika discussions. Many thanks go out to other Chalmerists, colleagues from elsewhere, friends, and of course my family.

- Magnus

Abstract

Breast Cancer (BCa) is the most common cancer for women in the Western world, and the American National Institutes of Health estimates that approximately 12.3% of American women will be diagnosed with breast cancer at some point in their lifetime[1, 2]. Breast MRI (Magnetic Resonance Imaging) is increasingly used clinically as an adjunct to traditional x-ray mammography and ultrasound for the detection and characterization of BCa. The interpretation of the large volume of image data acquired in a breast MRI exam is both a complex and time consuming task for the radiologist. Moreover, while the sensitivity of breast MRI to BCa is very high, its specificity is poor to moderate which results in large numbers of false positives. Computer automated/assisted detection/diagnosis (CAD) systems have been developed in an attempt to address these issues. However, a recent meta-study of breast MRI CAD systems concluded that they have little effect on the sensitivity and specificity of experienced radiologists[3]. Current literature suggests that performance gains may be achieved through 3D segmentation of suspicious lesions and the use of features (measurements) derived from multimodal MRI.

Typically a clinical breast MRI examination includes the acquisition of anatomical T1- and T2-weighted images, and a dynamic contrast-enhanced (DCE) sequence of T1-weighted images. In this thesis, a state-of-the-art method for creating spatially coherent clusters of similarly enhancing voxels, or voxels with similar signal intensity characteristics, was implemented and explored for use in a breast MRI CAD system. The method, Simple Linear Iterative Clustering (SLIC), generates an oversegmentation of the image into regions termed superpixels (2D) or supervoxels (3D). This method is an adapted k-means clustering approach that quickly and efficiently generates supervoxels/pixels by bounding the search area for similar voxels/pixels, and through the use of a simple distance metric[4, 5]. Here SLIC was used: i) to develop an algorithm to segment the breast-air boundary; and ii) to partition volumes of interest, corresponding to mass-like lesions, into supervoxels from which quantitative features describing the lesion are extracted. These features describe the contrast enhancement (perfusion) characteristics and diffusion characteristics of the lesion. Clinically acquired MRI data was used to evaluate both the segmentation method and the efficacy of the proposed features for discriminating between benign and malignant lesions.

Segmentations of the breast-air boundary were reviewed visually and found to adhere well to the boundary. Random Forest classification was used to estimate the classification performance of the proposed features, as well as to identify the most important subset of features. The results, based on a study of 77 subjects, show that the classifier is able to discriminate between benign and malignant lesions with an accuracy of 0.752 ± 0.055 (AUC$\pm$SE).

Collectively the results provide evidence that SLIC generated supervoxels are useful for both segmentation and classification in a CAD system. Further research is needed to investigate whether the combination of the proposed SLIC-based features and conventional features can improve the state-of-the-art in terms of sensitivity and specificity.

Keywords: Breast Cancer, Image Segmentation, Magnetic Resonance Imaging, Classification

Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

Breast Cancer (BCa) is the most common cancer for women in the western world, and the American National Institutes of Health (NIH) estimates that approximately 12.3% of American women will be diagnosed with breast cancer at some point in their lifetime [1, 2]. In 2013, the NIH estimated that there will be over 230 000 new cases of breast cancer, of which approximately 40 000 will be fatal[1].

The detection of breast cancer can be made in several ways, including but not limited to: X-ray mammography, physical examinations, ultrasound examinations, or magnetic resonance imaging (MRI). X-ray mammography is the standard modality used during screening procedures.

Recently, MRI has emerged as a complimentary imaging modality for the characterization of lesions found in the breast. Normally, an MRI examination of the breasts consists of several series of images of one or both of the breasts. T1- and T2-weighted images are acquired without contrast as well as a dynamic contrast-enhanced (DCE) series. The dynamic series normally involves the acquisition of one pre-contrast and several acquisitions after the patient is injected with a Gadolinium-containing contrast agent.

Radiologists use both the morphological features of the lesion and the contrast enhancement pattern to characterize the lesion in question. Using these two criteria, breast MRI has been shown to have a very high sensitivity to the detection of BCa, often exceeding 90%. However, the specificity of these studies is much lower, and has been reported between 67-72%[6, 3, 7] which yields a high number of false-positive results.

Worldwide, there are several screening programs for women who are of high risk for BCa, e.g. a family history of BCa. The American Cancer Society has recommended that women who have an estimated 20-25% lifetime risk for BCa have regular MRI screening[8]. Compounding the increase in screening for BCa, the increasing prevalence of this disease has lead to an influx of medical images that need to be acquired, processed, and analyzed by trained clinicians. Moreover, as MRI technology advances, the size of the data sets might increase further. The interpretation of these large data sets is complex, and time consuming. This has accelerated the research and development of computer aided/assisted detection/diagnosis systems (CAD) for use in breast MRI.

1.1 Breast Cancer

With over one million new cases worldwide each year, BCa is the most common malignancy in women and in total is responsible for one fifth of all female cancers [9, 10]. In the United States, the five-year survival rate across all BCa diagnoses is 89.2%[1].

There are several types of BCa, primarily classified by location and invasiveness.

Ductal Carcinoma in situ (DCIS): DCIS is considered a non-invasive BCa. When a patient has DCIS, it means that the cells that line the milk ducts in the breast have changed to appear as cancer cells. At this stage, the cancer cells have not spread through the walls of the ducts into the surrounding breast tissue. About one fifth of new BCa diagnoses will be a DCIS [2]. The prognosis is very good for this type of cancer and nearly all women diagnosed at this stage can be cured.

Invasive/Infiltrating Ductal Carcinoma (IDC): This is the most common type of BCa, accounting for 8 in 10 invasive breast cancers. IDC's originate in one of the milk ducts of the breast, breaking through the tissue wall to grow inside the fatty tissue of the breast. From there, it may be able to metastasize to other parts of the body via the circulatory or lymphatic systems.

Invasive/Infiltrating Lobular Carcinoma (ILC): ILC originates in the milk producing glands, the lobules. Similar to IDC, ILC can metastasize to other parts of the body using the circulatory or lymphatic systems. ILC is often more difficult to detect through mammography than an IDC.

Other less common types of breast cancer include: Inflammatory BCa, in which no single lump or lesion can be isolated, but the cancer cells are spread throughout the breast; Triple-Negative BCa, named as such because this cancer's cells lack oestrogen receptors, progesterone receptors, and do not have a specific protein marker on their surface; and Paget disease of the nipple, where the cancer starts in the breast ducts and spreads to the nipple. Triple-Negative BCa, as well as Paget disease of the nipple are both often associated with DCIS or IDC.

There are a wide variety of risk factors for BCa, including: age, ethnicity, age at menarche and menopause, family history, diet, body weight as well as environmental causes like exposure to radiation [9, 11]. The incidence of BCa increases with age, approximately doubling every ten years until menopause, when the rate of increase slows. BCa is also more prevalent in more developed, western nations.

1.2 Diagnosis of Breast Cancer

Diagnosis of BCa most often begins with the patient's medical history and a physical exam. Here, the physician will ask questions about the patient's symptoms, other health issues and run through potential risk factors. Often, patients see their physician after finding a lump in their breasts during a self-examination. After a physical examination, a clinician often determines that imaging studies are needed. Multiple imaging modalities can be used, including X-ray mammography, ultrasonography, and MR imaging. Many developed nations have large screening programmes designed with the goal of early detection

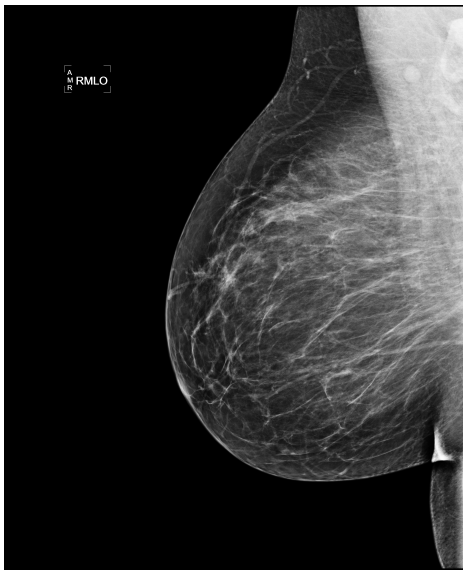


Figure 1.2.1: Sample mammogram.
Reproduced from:[13]

in mind. These programmes are based on X-ray mammography.

In X-ray mammography, a two-dimensional X-ray image is captured of the breast(s), usually using 2 orthogonal views. See Figure 1.2.1 for an example. X-ray mammography is used most often as it is an inexpensive, and very fast imaging study, and can often immediately determine that an abnormality is not worrisome. It is conventionally used to both assess symptomatic patients and for screening purposes. X-ray mammography has several drawbacks: when performed optimally, the sensitivity of X-ray mammography is between 69-90%[12]; as a radiographic technique, patients are exposed to dangerous ionizing radiation when they undergo an investigation; and, given that X-ray mammography is a 2D projection-based method, the resulting image can be blurred and is difficult to read. It also has poor soft-tissue contrast and cannot be used to measure vascularity of the breasts. Often, a follow-up biopsy is recommended to confirm a diagnosis. Images from women who have particularly dense breasts can be difficult to read.

In MRI, the hydrogen nucleus, or proton is used because of its high concentration in biological tissues like fat and water. MR imaging uses magnetic fields to excite the protons and record how they respond to specific magnetic field pulse sequences. MRI can be used in a variety of acquisition modes, creating multiple contrast patterns, such as T1- and T2-weighted, diffusion weighted (DW), and can also be used to monitor the uptake and distribution of contrast material. Dynamic Contrast Enhanced (DCE) MR images are the most routinely used MRI technique for the assessment of breast lesions [12]. Under guidance from the American Cancer Society, breast screening is recommended for patients in specific risk groups such as those with a family history of breast

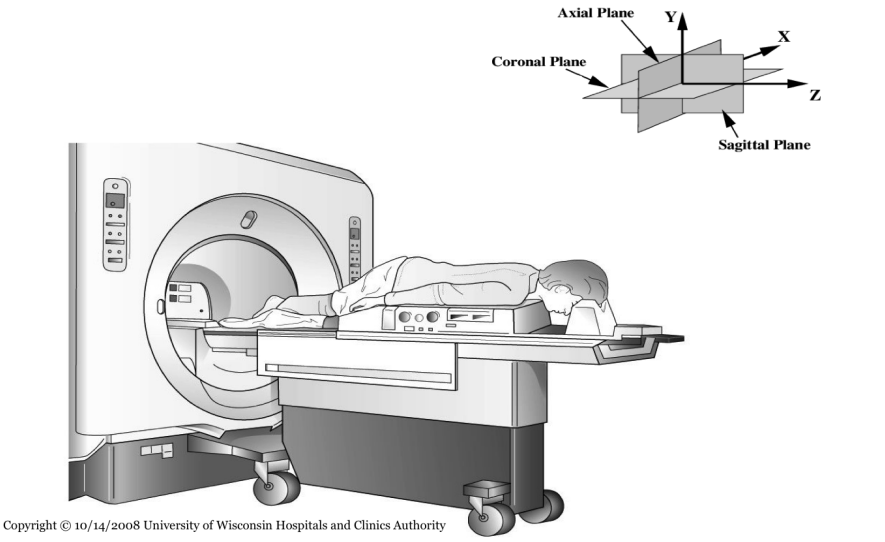


Figure 1.2.2: Schematic of Breast MRI acquisition

cancer or who face a lifetime risk of 20-25%[8]. MRI has higher sensitivity and can find smaller tumours than conventional mammography [12, 8]. DCE-MRI uses volumes acquired at several time instances, before and after the injection of a Gadolinium-containing contrast agent, to assess the contrast uptake and washout patterns of the tissue. A typical DCE-MRI exam will include one pre-contrast volume followed by four to six contrast-enhanced volumes. A schematic view of a patient lying in an MRI scanner showing the coils used for breast imaging is shown in Figure 1.2.2.

In many practices, the radiologist reports findings using the American College of Radiologist's Breast Imaging and Reporting Data Systems (ACR BI-RADS)[14], which provides a standardized way for a radiologist to report findings in the breast MRI exam[1]. The assessment of the lesion is provided on a 0-6 scale, where an increasing value suggests a more suspicious lesion. Category 1 reports a negative finding, Category 2 reports a benign finding. Categories 3 through 5 report that the findings are increasingly suggestive of malignancy. Category 6 is used to report a known cancer. The assessment of the lesion is based on several categories, including:

- **Shape** (for a mass-like enhancing lesion)
  - Round
  - Oval
  - Irregular or Lobular
- **Enhancement Properties**
  - Homogeneity

[1] BI-RADS has not been adopted in every country

  - Heterogeneity
  - Rim Enhancement

- **Margin Features**
  - Smooth
  - Irregular or Spiculated

- **Other**
  - Nipple Involvement
  - Skin thickening
  - Edema

- **Enhancement Characteristics over time**

1.3 Computer Assisted Diagnosis/Detection

Computer assisted diagnosis/detection (CAD) systems were first developed to identify suspect lesions in the breast volume and bring them to the attention of the radiologist, but thereafter the primary aim shifted to distinguishing between benign lesions and malignant lesions [3, 15, 16, 12]. In this arrangement, the program aims to help decrease the number of false-negative readings by radiologists as opposed to the former, where CAD is used as a second-reader[2]. However, most suspicious lesions flagged by the CAD program were already detected by the radiologist. This turned the attention of CAD developers to the task of assisting radiologists determine if a suspect lesion is benign or malignant.

Successful implementations of CAD software would improve the accuracy of breast MRI by cutting the false-positive rate (and possibly increasing the sensitivity) and reduce the amount of time clinicians and technicians need to spend on processing and interpreting the images. State-of-the-art CAD systems have very high sensitivities and negative predictive values[3] for non-calcified breast lesions[3].

A recent meta-study of ten studies on how CAD affects radiologists both experienced and otherwise suggests that those who have little experience can benefit from the use of CAD systems. However, experienced radiologists saw little or no benefit [3]. The meta-study reported a pooled sensitivity of 89% and a pooled specificity of 81% for both radiologists and residents assessing breast lesions with the use of a CAD system, versus reported sensitivities of above 90% and specificities ranging between 67%-72% without CAD.

CAD systems perform many of the preprocessing and analysis functions that a radiologist or the MR technician would normally have to do themselves. The software also automates the generation of kinetic curves[4] for the breast tissue,

[2] In a second-reader scenario, the software is used as a second radiologist would be - to analyze the images, and ensure all lesions are found.

[3] The Negative Predictive Value (NPV), is the ratio of the number of true negatives to the total number of negative classifications in a test.

[4] These kinetic curves show the relative signal intensity for the lesion in question before, and after the injection of contrast material. The lesion's kinetics describe the rate at which contrast flows in and out of the tissue.

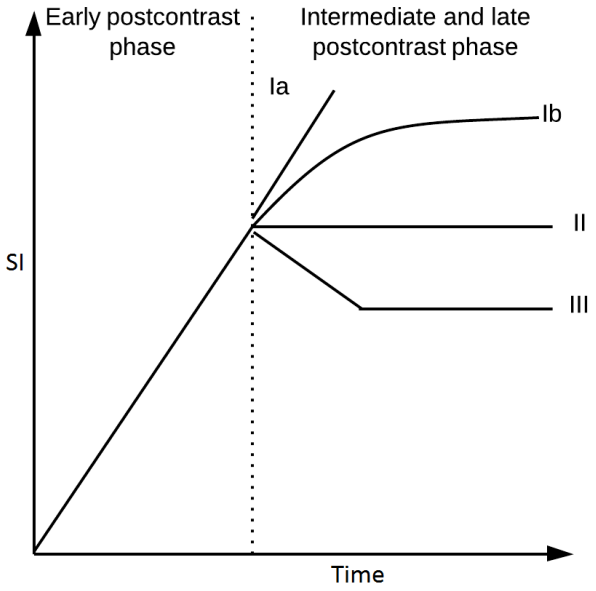


Figure 1.3.1: Schematic of time-signal intensity (SI) curves. Type I corresponds to a straight line (Ia) or curved (Ib) and reflects an enhancement which continues over the whole study. Type II is a plateau curve. Type III is a washout pattern. A type III pattern is a strong indicator of malignancy [12].

and provides a visualization that displays the patterns of contrast enhancement (See Figure 1.3.1) across a series of images [7, 3]. The results of this assessment are often displayed to the user as a colour-coded map based on the uptake strength [7].

State-of-the-art CAD software takes into account many features of the suspect lesion, including it's morphology, enhancement characteristics, among others [3]. As many analysis procedures performed by radiologists are subjective (e.g. assessing the morphology of the lesion), there exists high inter-observer variability that can result in differing diagnostic performance[17]. CAD systems remove this variability, processing every image in the same manner. A wide variety of features anecdotally described by radiologists have been collected in [18], and work is being done to translate these qualitative descriptions to quantitative measures [15, 19].

Several automated analysis methods are being developed to quantify the uptake (or wash-in), and wash-out of the contrast agent, and to divide the voxels within a lesion into spatially contiguous clusters that have similar enhancement patterns. Using spatially contiguous clusters inside the lesion, analysis of the temporal shape of the enhancement pattern can also be analyzed. There are a wide range of parametric models used in literature to describe enhancement patterns, and several are used in this thesis[6, 20, 21, 22, 23, 24].

A typical work flow for a breast CAD system would include first high-level

image enhancement techniques, such as contrast adjustments, bias field corrections, or noise reduction algorithms. Next, areas that are not important in the image would be removed. This may include the removal of air, and the removal of the chest-cavity from the volume. After this large-scale segmentation is complete, the volume would contain only the voxels corresponding to the breast tissue. Next, a lesion would be identified automatically, most likely from the T2-weighted and dynamic volumes. Then, the lesion itself is segmented into spatially coherent clusters. Next, information is gathered from these clusters which could relate to their morphology, enhancement pattern, or diffusion characteristics. Finally, the information gathered from the clusters is pushed to a classification system which analyzes the information and presents a determination as to whether or not the lesion is benign or malignant.

In this thesis, a new method for creating these spatially coherent clusters of similarly enhancing voxels, or voxels with similar signal intensity characteristics, was implemented and explored for use in a CAD system. The method is based on Simple Linear Iterative Clustering (SLIC), which is a new method used to generate an oversegmentation[5] of regions with similar characteristics [4, 5]. Superpixels, in the two-dimensional case, and supervoxels, in the three-dimensional case, are the result of the oversegmentation process[25]. This method is an adapted k-means clustering approach that quickly and efficiently generates supervoxels by bounding the search area for similar voxels, and through the use of a simple distance metric[4, 5].

1.4 Data and Software Used

The MRI data used in this research was acquired from routine clinical examinations of women performed by Queensland X-ray in Australia. The examinations were used for screening, investigative purposes, or as follow up studies after surgery. The images were acquired using a 1.5T scanner. In each case, the reporting radiologist identified at least one suspiciously enhancing lesion, and pathology for each lesion was subsequently confirmed by cyto- or histopathology. Lesions were individually biopsied under either MRI or ultrasound guidance.

Each case consists of T1- and T2-weighted anatomical volumes, a two diffusion weighted volumes, and five dynamic contrast enhanced volumes. Volumes were spatially co-registered and bias field correction performed. The T1- and T2- weighted registered images used in this work normally have size 512×512 pixels, with on average 150 images in the stack. The detailed acquisition, and registration protocol can be found in [15, 19]. Lesions were delineated manually by a radiographer with 12 years experience in breast MRI, guided by the findings of the original reporting radiologist. Manual delineations were performed on the subtraction volume of the first post-contrast volume from the pre-contrast volume using a region growing tool in OsiriX. The radiographer also classified the lesion as mass-like or not, guided by the American College of Radiology BI-RADS lexicon. All lesions have a classification of BI-RADS 2+.

A subset of the database of subjects was selected for use in lesion segmentation and feature selection experiments. To be included in the subset which was

[5] An oversegmentation describes a segmentation which has more segments than true objects in the image. A perfect segmentation would have the same number of segments and objects, and an undersegmentation would therefore have fewer segments than objects.

used:

1. The suspicious tissue must have been classified as mass-like;
2. At least 4 DCE-MRI volumes were available, where the first volume is before the administration of contrast material; and
3. 2 DW volumes were available, or, an apparent diffusion coefficient map was available as a substitute.

After these requirements were used as filters, 77 acceptable subjects remained. Of those subjects, 53 were diagnosed as having malignant lesions, while 24 were diagnosed as having benign lesions. The lesion pathologies are shown in Table 1.1.

Table 1.1: Pathologies of Selected Lesions

Pathology	Count	Description
Benign	4	Fibrocystic change
	9	Fibroadenoma
	11	Other
Total Benign	24	
Malignant	14	Ductal carcinoma in situ
	20	Invasive ductal carcinoma
	3	Invasive lobular carcinoma
	16	Other
Total Malignant	53	
Total Mass-like	77	

Algorithm development, testing was performed in the MATLAB[26] programming environment. The open source OsiriX (www.osirix-viewer.com) viewer was used for visualization and exploration of the MRI volumes. The classification algorithms were developed and executed using R[27].

1.5 Aim and Objectives

The aim of this research was to implement and explore the SLIC supervoxel algorithm for use in a computer aided diagnosis system for breast cancer. To that end, the research had the following specific objectives:

1. To implement SLIC for the generation of supervoxels in breast MRI data;
2. To develop and evaluate a supervoxel-based method to segment the breast-air boundary; and
3. To develop and evaluate supervoxel-based features for classifying breast lesions as benign or malignant.

1.6 Scope

The focus of this research was restricted to the development of algorithms based on supervoxels, generated using the SLIC algorithm, for use in a breast MRI CAD system. In particular the research was restricted to the problem of segmenting the breast air boundary and the extraction of lesion features.

1.7 Structure of Thesis

This chapter has:

- Provided background on BCa, described several presentations of BCa, and noted several risk factors.
- Discussed the various methods used to diagnose BCa, noting their respective advantages and disadvantages.
- Discussed the rationale for the use of CAD systems in Breast MRI.
- Described the data and software used in the thesis.
- Presented the aim, objectives, and scope of this thesis.

The remainder of the thesis is organized as follows:

Chapter 2 This chapter addresses the first objective of the thesis: SLIC segmentation for the generation of supervoxels in breast MRI.

Chapter 3 This chapter addresses the second objective of the thesis: to develop and evaluate a supervoxel-based method to segment the breast-air boundary.

Chapter 4 This chapter presents the method used for generating supervoxels within a volume-of-interest, and in combination with chapters 5 and 6 addresses the third objective of the thesis: to develop and evaluate supervoxel-based features for classifying breast lesions as benign or malignant.

Chapter 5 This chapter builds on the previous to address the third objective in the thesis. Here, several features quantitative features are derived from the supervoxels generated within the volume-of-interest.

Chapter 6 This chapter describes classification experiments to determine the discriminatory performance of the features proposed in the previous chapter.

Chapter 7 This final chapter reviews the work that has been presented in this thesis, and summarizes the major contributions and findings. The limitations of the research undertaken, and potential future research is also discussed.

Chapter 2

SLIC Segmentation

This chapter addresses the first objective of this thesis: to implement SLIC for the generation of supervoxels in breast MRI. Superpixels, and their 3D extension termed supervoxels, are increasingly popular in computer vision and have been used in a variety of applications such as depth estimation, image segmentation, skeletonization, and object localization [4, 28, 29, 25]. They can also be used to reduce the complexity of image processing tasks by capturing redundancy. If superpixels or supervoxels are to be used in a complexity reduction scenario, their creation must be simple to tune and at low computational cost. The remainder of this chapter is organized as follows. In the next section the rationale for choosing the SLIC supervoxel algorithm is discussed. Section 2.2 provides a description of the SLIC algorithm. Section 2.3 discusses the author's implementation in MATLAB. Section 2.4 shows example segmentations with respect to different parameter settings.

2.1 Rationale

Several algorithms can be used to create superpixels/voxels, and they can be broadly classified into two different classes: graph-based, and gradient-ascent-based. Simple Linear Iterative Clustering (SLIC) falls into the latter category and is described as an adaptation of the k-means approach.

There are two main distinctions between SLIC and k-means. The first is that the number of distance calculations during the optimization phase is significantly reduced by reducing the search region to a space proportional to the expected supervoxel size (See Figure 2.2.1). The second is that a weighted distance measure is used which combines the intensity information and the spatial proximity to provide control over the size and compactness of superpixels/voxels.

Not all of the algorithms described in literature can be extended directly into 3D, nor provide control over the number of superpixels/voxels. Many have a large numbers of parameters that need to be tuned. SLIC, on the other hand, has only one tuning parameter; control over the number of superpixels/voxels and can be extended into 3D. It was for these reasons that SLIC was chosen for this thesis. As this application deals with 3D MRI volumes, the supervoxel case will be discussed from this point forward.

Most commonly, SLIC generates superpixels based on their colour similarity

and their spatial proximity in the 2D image plane. Herein, it is adapted to generate supervoxels based on their MRI intensities from one or multiple MRI modalities and their spatial distance in three dimensional space. Other popular superpixel methods can potentially be extended into 3D, including Quickshift [30], Turbopixels [31], and graph-based algorithms[32]. However, these methods have been shown to produce lower quality segmentations in 2D, can be orders of magnitude slower in runtime, and require more memory [5].

Instead of generating supervoxels for the entire volume, it is also possible to create superpixels on each level of the image stack. This approach was compared in [33] to the true-3D supervoxel approach in brain MR image analysis. They found that while superpixels show better boundary adherence at small levels of complexity reduction, supervoxels show much better boundary adherence when the reduction in complexity[1] is higher. They showed that supervoxels achieved a higher reduction in complexity (approximately 200x) than superpixels (approximately 75x) while maintaining the same levels of boundary adherence [33]. SLIC has $O(N)$ complexity [4, 5, 34], where N is the number of voxels, compared to a conventional k-means approach which has complexity on the order of $O(kN)$ where k is the number of clusters.

2.2 SLIC

The SLIC algorithm is straight forward, and can be broken down into several stages.

The first phase is the initialization phase. The creation of cluster centres begins with an initialization of k clusters equally spaced throughout the volume, where k represents the desired approximate number of equally sized supervoxels. Taking I_{ni} to be the MRI signal intensity for the cluster centre i in modality n, and x,y,z, being the spatial coordinates, the cluster centres $C_i = \begin{bmatrix} I_{1i} & I_{2i} & \cdots & I_{ni} & x_i & y_i & z_i \end{bmatrix}^T$ are sampled on a lattice spaced S voxels apart, where:

$$S = \sqrt[3]{N/k} \tag{2.2.1}$$

and N is the total number of voxels in the volume. Centres are moved to the lowest gradient position within their local $3 \times 3 \times 3$ neighbourhood to avoid centring a supervoxel on an edge and to reduce the chance that the seed voxel is a noisy voxel.

The next phase is the assignment phase, where each voxel i is associated with the nearest cluster centre whose search region overlaps it. The distance measure D determines the nearest cluster centre for each voxel. Given that the expected spatial extent of a supervoxel is a volume of approximately $S \times S \times S$, the search for similar voxels is performed in a region of $2S \times 2S \times 2S$ around the supervoxel centre.

SLIC combines the intensity distance (d_I) and spatial distance (d_s) into a single metric, D. In order to combine the two distances into this single metric it is necessary to normalize both the spatial and intensity proximities by

[1] Here, complexity reduction refers to the number of elements in the image(volume) that need to be analyzed. This can also be used as an estimate of the average size of the superpixel/voxel.

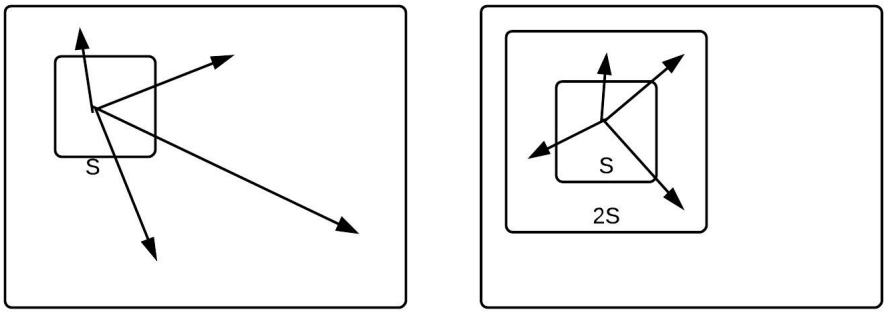


Figure 2.2.1: Diagram of SLIC search region inside an image.
A standard k-means approach searches the entire image for similar voxels (left), whereas SLIC searches a smaller, bounded region (right). This increases the computational efficiency of SLIC. This idea is similarly extended into the third dimension.

their respective maximum distances within a cluster N_s and N_I. Therefore the normalized metric D' is formed by combining the spatial and intensity distances:

$$d_I = \sqrt{(I_{1i} - I_{1j})^2 + (I_{2i} - I_{2j})^2 + \dots (I_{ni} - I_{nj})^2}, \tag{2.2.2}$$

$$d_s = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + \alpha(z_i - z_j)^2}, \tag{2.2.3}$$

$$D' = \sqrt{(\frac{d_I}{N_I})^2 + (\frac{d_s}{N_s})^2} \tag{2.2.4}$$

The scalar factor α is used to adjust for anisotropic voxel dimensions in the volume, and is defined:

$$\alpha = \frac{l_{x,y}}{l_z}, \tag{2.2.5}$$

where $l_{x,y}$ is the length and width in the xy plane, and l_z is the height of the voxel.

The maximum spatial distance expected within a given supervoxel corresponds to the sampling interval,

$$N_s = S = \sqrt[3]{N/k}. \tag{2.2.6}$$

Determining the maximum intensity distance is not as straightforward however, as intensity distances vary supervoxel to supervoxel and volume to volume. This problem is solved by setting the normalized intensity distance to a constant m so that the distance equation becomes:

$$D' = \sqrt{(\frac{d_I}{m})^2 + (\frac{d_s}{N_s})^2}, \tag{2.2.7}$$

which can be simplified to:

$$D = \sqrt{d_I^2 + \frac{d_s}{S}^2 m^2}. \tag{2.2.8}$$

Creating the distance measure in this way allows for m to be used to weigh the relative importance of the intensity similarity and the spatial proximity. When m is large, spatial proximity is weighed more heavily and this produces compact supervoxels. Conversely, when m is small the supervoxels created follow image boundaries closely and have more irregular sizes and shapes. S and m can be thought of as the average expected spatial and intensity distances inside a supervoxel, respectively[34]. Compact supervoxels can be desirable becasue they more often correspond to a lattice structure in the volume, and their boundaries are simpler which leads to more regular neighbourhood relationships.

The compactness parameter m can also be described in relation to the intensity range in the volume:

$$m = range_{\%} \cdot (\max(I) - \min(I)). \tag{2.2.9}$$

This definition yields a more intuitive procedure for the user as now the user needs to input a target average intensity range for the supervoxels, and this is scaled to each particular volume.

After each of the voxels have been assigned to a cluster, the update phase begins. The cluster centre positions are updated to the mean $\begin{bmatrix} I_{1i} & I_{2i} & \cdots & I_{ni} & x_i & y_i & z_i \end{bmatrix}^T$ vector of all of the voxels in the cluster.

The assignment and update phases are repeated iteratively until the L2 norm error between the new cluster centre locations and the old locations converge. Alternatively, a fixed number of iterations can be used.

The final phase is a post-processing phases to enforce connectivity by reassigning any disjoint voxels to nearby supervoxels, and ensuring that all voxels within a label are connected components within 26-neighbour (neighbours are defined through all faces, edges, and corners) connectivity. In this process, a minimum supervoxel size can also be enforced. Other connectivity schemes can also be used, such as 6-connectivity (all of the faces on the voxel element), or 18-connectivity (all of the faces and edges).

The resulting segmentations can be described using the average number of voxels per supervoxel, which correspond to the reduction in complexity of the volume:

$$C = N_{avg} = \frac{N}{k}. \tag{2.2.10}$$

An overview of the algorithm can be seen in Algorithm 2.1.

2.3 Implementation in MATLAB

As there were no publicly available implementations of the SLIC algorithm for use in breast MRI, the SLIC algorithm was implemented in MATLAB by the author. The implementation includes methods for reading in DICOM[2] image stacks, the SLIC algorithm itself, a connected-components method to enforce supervoxel connectivity, methods to create masks used for visualization, and a graphical user interface to allow users to explore the images. The implementation also allows for multiple MRI modalities and provides the ability to restrict clustering to a specific volume of interest. Where possible, the implementation

[2]Digital Imaging and Communications in Medicine (DICOM) is the standard file format for medical images

Algorithm 2.1 Simple Linear Iterative Clustering

Input: Breast MRI volume, approximate desired number of supervoxels (k), compactness parameter (m)

Initialize cluster centres $C_k = \begin{bmatrix} I_{1k} & I_{2k} & \cdots & I_{nk} & x_k & y_k & z_k \end{bmatrix}^T$ by sampling at regular grid intervals S

Perturb cluster centres to lowest gradient position in $3 \times 3 \times 3$ region

Set labels $l(i) = -1$ for each voxel i

Set distance $d(i) = \infty$ for each voxel i

Repeat

    for each cluster centre C_k

        for each voxel i in $2S \times 2S \times 2S$ region around C_k

        Calculate distance D between C_k and i

            if $D < d(i)$ then

                set $d(i) = D$

                set $l(i) = k$

            end

        end

    end

    Compute new cluster centres

    Compute residual error

until Error $\leq$ Threshold or Number of Iterations reached

Enforce Connectivity

Output: Labelled volume

was written in a parallel manner, to take advantage of modern multi-core processors. The implementation package is available on request from the author.

2.4 Example Segmentations

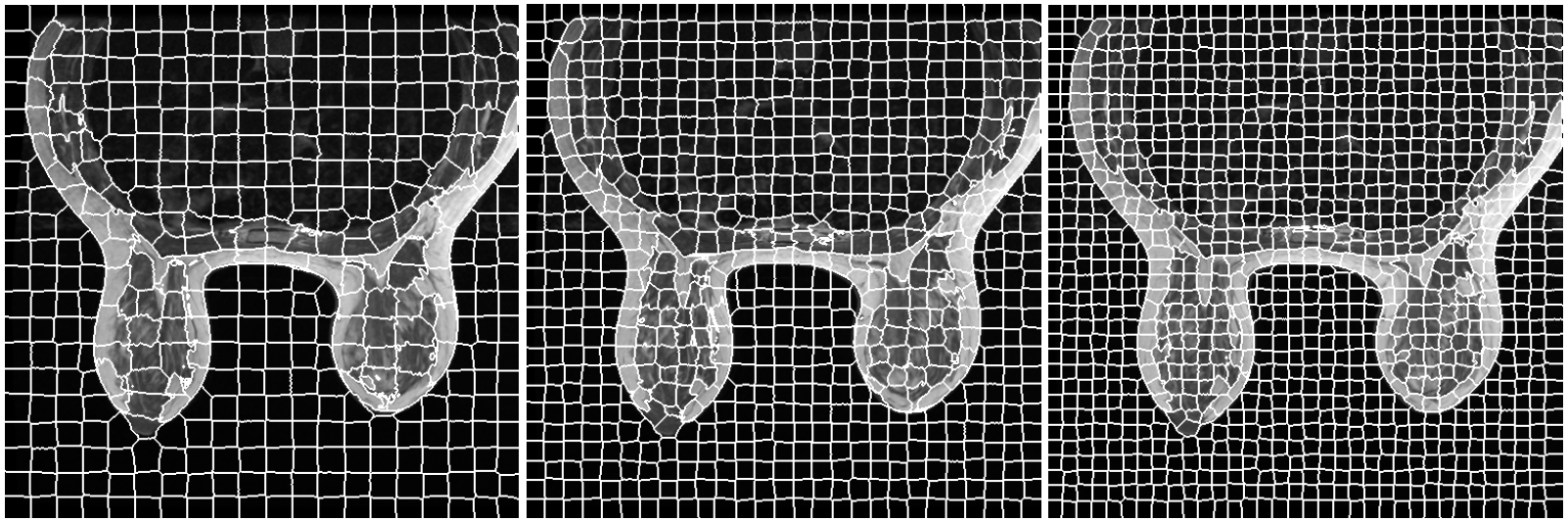
In this section the output of the SLIC supervoxel algorithm will be shown.

Figure 2.4.1 shows the supervoxel output for a range of supervoxel cluster sizes.

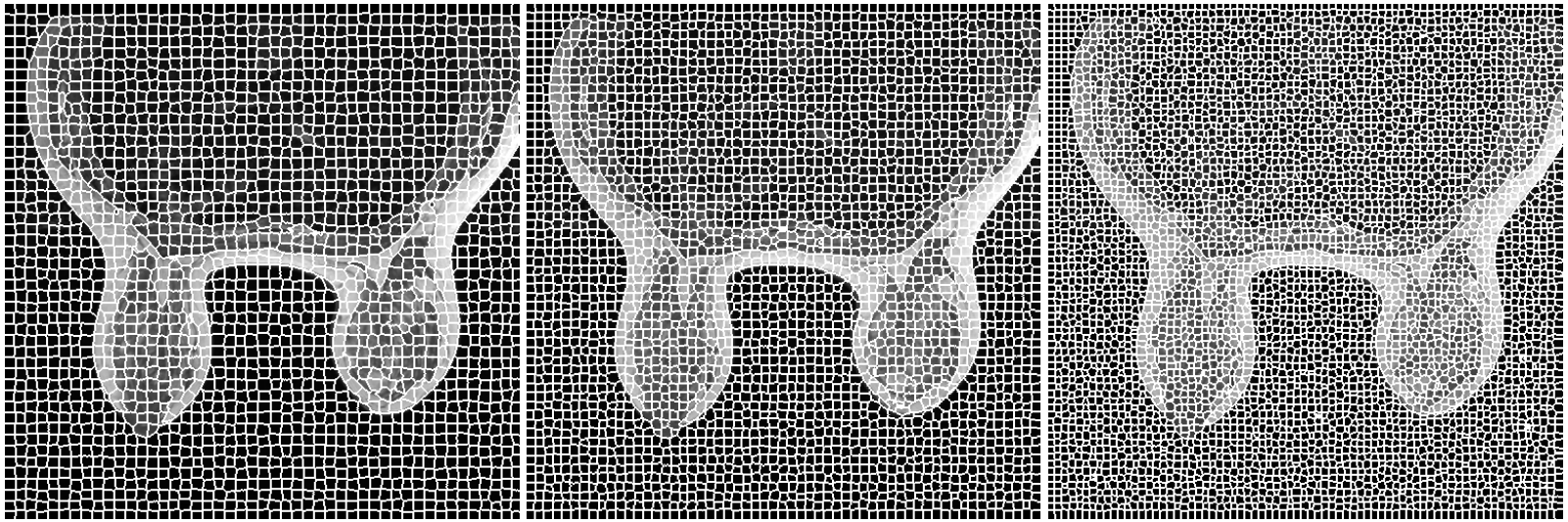
As the complexity reduction decreases the homogeneity of the clusters increases as the supervoxel has aggregated a much smaller number of more similar voxels. At each level the supervoxels continue to adhere to image boundaries, and in areas of the image where there is little to no variance in the intensity the supervoxels form a more regular lattice.

Figure 2.4.2 shows the supervoxel output for a constant complexity reduction level, but for a range of compactness parameters.

It can be seen that as the compactness parameter rises, each supervoxel is allowed to have a greater range of intensity values within itself. It follows that if an image boundary shows an intensity difference of larger than this expected range, it is more likely to be excluded from that superpixel (although this is also dependent on the spatial distance from the supervoxel centre). For example, in Figure 2.4.2a, the supervoxel boundaries are much more complex as they attempt to follow the internal boundaries of the breast tissue, in contrast with Figure 2.4.2d the supervoxels have very regular shapes, even inside the breasts.

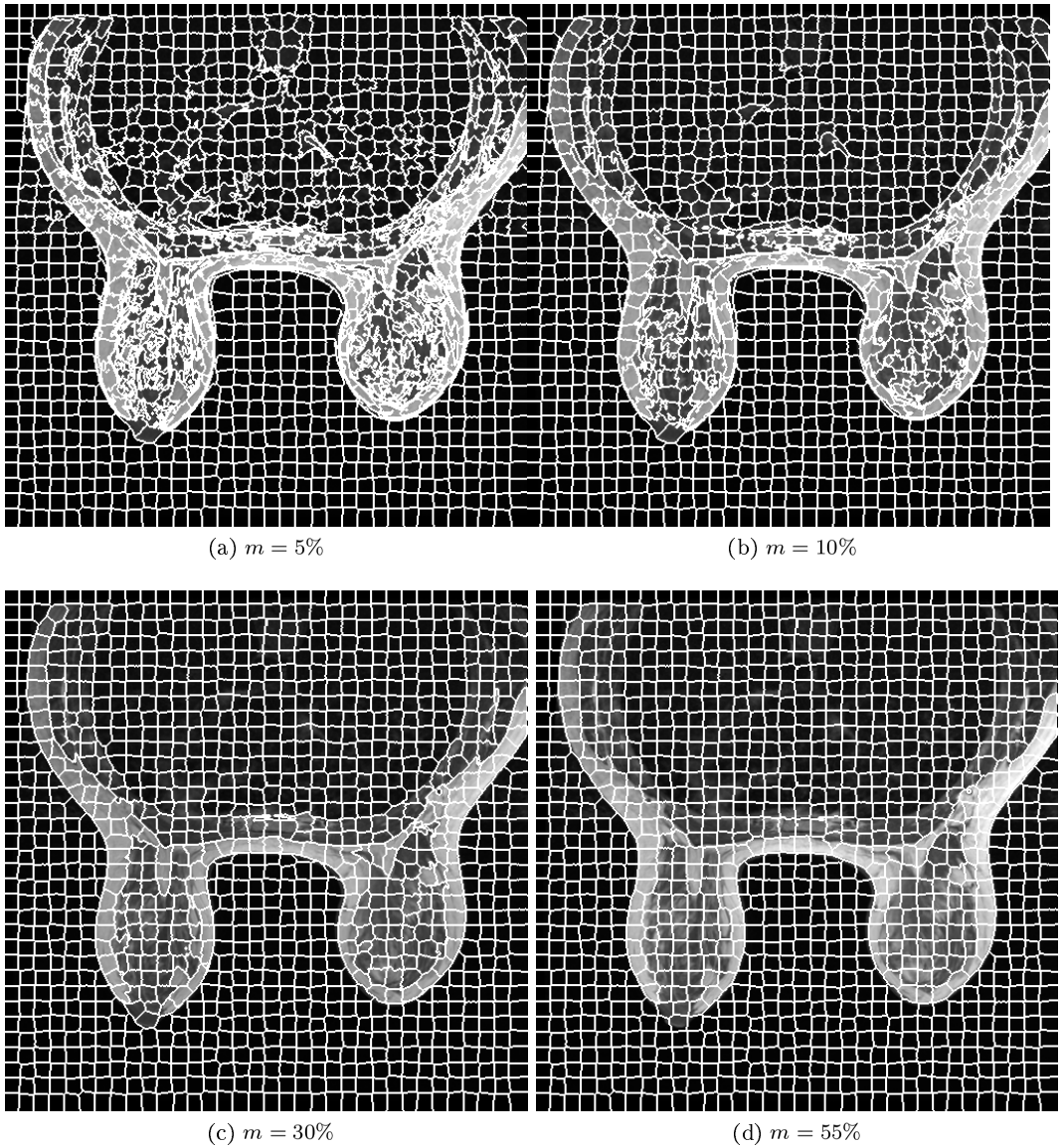


(a) $k = 2\ 400$ (b) $k = 5\ 800$ (c) $k = 12\ 000$



(d) $k = 28\ 000$ (e) $k = 52\ 000$ (f) $k = 110\ 000$

Figure 2.4.1: SLIC supervoxels at various sizes.
Segmentations from the same T1-weighted volume, at the same slice. Compactness parameter $m = 500$, or approximately 35% of the intensity range. With segmentations using these total numbers of supervoxels, the resulting complexity reductions are approximately: 16000, 6700, 3200, 1350, 735, and 350 respectively.



(a) $m = 5\%$ (b) $m = 10\%$

(c) $m = 30\%$ (d) $m = 55\%$

Figure 2.4.2: SLIC supervoxels with varying levels of compactness.
Segmentations from the same T1-weighted volume, at the same slice. $k = 12000$, yielding complexity reduction $C = 3200$.

Chapter 3

Segmentation of the Breast-Air Boundary

This chapter addresses the second objective of this thesis: to develop and evaluate a supervoxel-based method to segment the breast-air boundary.

Preliminary, large-scale image processing tasks need to be carried out on the images before further analysis can be performed in a CAD system. As breast MRI volumes are often acquired in the axial plane, they contain large areas that are of no use to the radiologist with respect to the diagnosis, namely the background of air and the chest cavity. It is important to remove these areas to be sure that any subsequent analysis tasks are performed solely within the breasts as well as reduce computation time. A typical image in the axial plane is shown in Figure 3.0.1, and the image includes the large background section of air, and the chest cavity, in which the heart is visible. Isolation of the breast tissue involves both the segmentation of the breast-air boundary and of the chest wall. This chapter considers only the first segmentation problem. The remainder of this chapter is organized as follows. In the next section the breast-air boundary segmentation algorithm of Hayton [21, 35] is described. It serves as a benchmark for the proposed SLIC supervoxel algorithm presented in Section 3.2.

3.1 Hayton's Method

Several methods exist to segment the breast-air boundary in digital mammography or MRI volumes, for example: seeded region growing [36], gradient analysis, active contours (snakes) [37], and dynamic programming [21, 35]. The ideal method would be fully automatic, fast, and reliable[1]. One approach that largely fits these demands is the method of Hayton [21] that is based on a live-wire boundary extraction algorithm that utilizes a graph search algorithm, a dynamic programming technique, to find the path of lowest cost[2] between two

[1] Reliable, in the sense that repeated user intervention is not needed for tuning or troubleshooting.

[2] Cost is most often based on image gradients, where traversing a path that includes large changes in gray-level is 'expensive'.

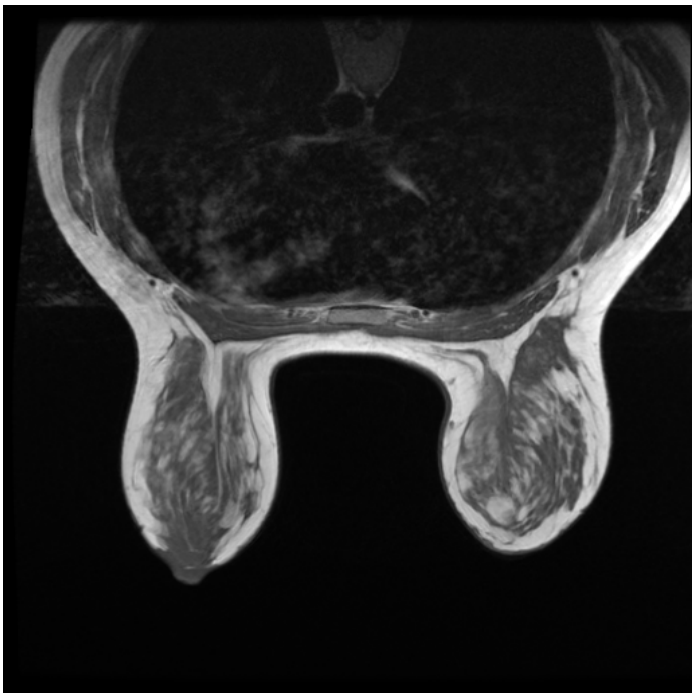


Figure 3.0.1: Typical axial image from a T1-weighted breast MRI volume

points on an image (this path was then used to make a computer-assisted segmentation tool). Hayton developed this algorithm further to run automatically, finding the lowest cost path between five automatically determined points on the breast profile (lower extremities on each breast, the outer edges of the chest, and a point between the breasts, see Figure 3.1.1a). Hayton's algorithm is shown in Algorithm 3.1 and is briefly presented below, as well as examples of its output.

First, using a simple thresholding operation, and prior knowledge of the orientation of the breast images, a contour of the breasts is created by searching vertically for the first strong edge, which should correspond to the breast-air boundary. This contour is not a robust edge contour, as it cannot wrap around the breast to capture the full shape of the breast, but it does provide an estimate for the lowest point of each breast, the uppermost point between the breasts, and the outer bounds. These points will be used as seed and target points while searching for the optimum path. An example contour with the target points is shown in Figure 3.1.1a. The graph search algorithm used is initialized with a local cost map from which the algorithm determines the path with the lowest total cost. Hayton uses an exponential local cost function which yields low values at tissue boundaries, and tends to its maximum of 1 in regions of uniform intensity. It is defined as:

$$C(x,y) = \exp(-\frac{\lambda|G(x,y)|}{\max_{x,y}|G(x,y)|}) \tag{3.1.1}$$

where λ is a scaling factor set experimentally, and $G(x,y)$ is the image gradient.

Next, the algorithm computes a cumulative cost map by starting at one of the target nodes (the predetermined points from the breast contour) and expanding

outwards. From the local cost map $C(x, y)$ the target point is expanded by summing the local costs of neighbouring pixels with the expanding pixel's total cost. The next lowest total cost pixel is then expanded to its neighbouring pixels, and so on throughout the image. When a node N is expanded, its neighbours consider the cost of the path back to the target point via N, and since N is being expanded, it must have the lowest cumulative cost, and therefore its path back to the target point must be optimal. Neighbouring pixels update their total costs, and point to N in the path map if going through N reduces their total costs. Therefore, from path map the minimum cumulative cost path from the seed point to the target point can be found by following the cost map, always selecting the neighbour node with the lowest cost as it points the route to the target. An example cost map is shown in 3.1.1b. This can be thought of as a concave surface, where the path rolls down the surface into the target node along strong edges of the image.

Using the five predetermined points, the method runs four times, twice from the outer boundaries to the lower breast extremities for each breast, and twice more from the centre-point between the breasts to the extremities on each breast. The cost map is generated separately for each breast. Once each section of the contour has been found, they are connected and the region above the contour corresponding to the breasts, is filled in to create a binary image mask. An example of the output binary mask is shown in Figure 3.1.1c. This procedure is then repeated on each image in the volume. The algorithm is summarized in Algorithm 3.1.

Two issues present themselves when using the Hayton algorithm: first, that computing two full cumulative cost maps per image in the volume is not efficient; and, that the output can still be somewhat unstable due to partial volume effects[3] or regions of low intensity. The latter can be solved to an extent by tuning the threshold for each volume to be segmented, ensuring that the threshold selected is high enough to remove noise, but low enough to capture areas on the edge which have been affected by partial volume effects. The former however is not as readily dealt with. There exist methods to speed up the algorithm, such as limiting the generation of the cumulative cost map to a bounded area[4], exploiting the similarities between slices, or retooling the algorithm to run in parallel[5].

Below are several examples of problematic output from the algorithm. In Figure 3.1.2a the path has wandered away from the breast-air boundary, following instead an area of lower signal intensity inside the breast. In Figure 3.1.2b the path has similarly wandered away from the breast-air boundary, although in this more extreme case the resulting segmentation failed to capture the breast.

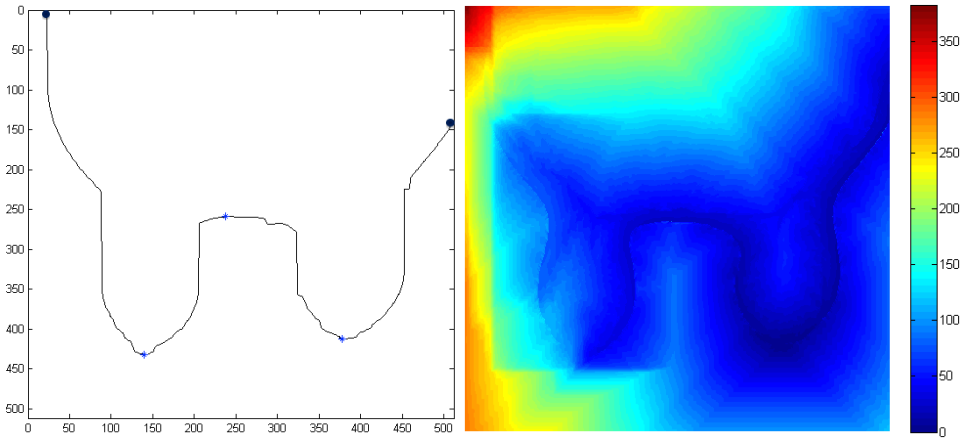
3.2 Proposed Method

In this section, a method for segmenting the breast-air boundary using supervoxels is proposed. In this method, the full breast-air surface is segmented

[3] Near regions of large gradients in the image, the low intensity side of the gradient often gets boosted to a stronger intensity. In particular, this is called spillover.

[4] For example, not generating the left breast cumulative cost map past the centre-point in the x direction or further than the outer boundary seed point in the y direction.

[5] As each slice in the volume can be treated independently using the Hayton algorithm, this algorithm is easily parallelizable, and significant speed increases can be found.



(a) Edge Contour of the Breasts

(b) Cumulative cost map for right breast

(c) Output Binary Mask

Figure 3.1.1: Output from various stages of the Hayton breast-air boundary segmentation algorithm.

Algorithm 3.1 Hayton Breast-Air Segmentation

Input: Image stack, threshold T

For each image in the volume

Construct Estimates for breast edge locations

For each x position, find lowest y position that has gradient magnitude $>$ threshold T to create the contour

Filter the contour to reject outliers and smooth it using median filter

Find Extremity points corresponding to outer bounds, the breast extremities, and centre-point

Compute local cost map $C(x, y) = \exp(-\frac{\lambda|G(x,y)|}{max_{x,y}|G(x,y)|})$

Between each extremity

Initialize sorted list of pixels S to be expanded from the target point

Initialize Cumulative cost table C_t with $C_t = 0$ at the target point, otherwise $C_t = \infty$

Repeat until all pixels have been considered

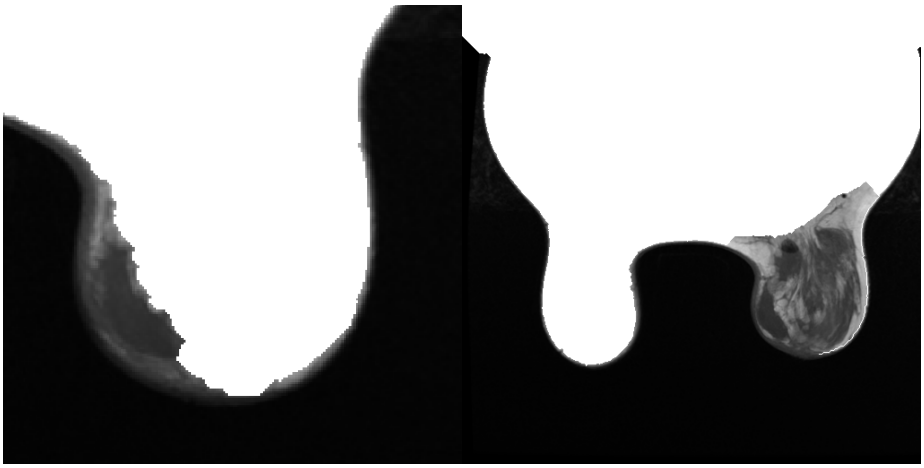
Expand Cumulative cost table outwards from target point

Trace the path of lowest cost between seed point and target

Connect each segment of the full path

Fill above the full path to create binary mask

Output: Stack of binary masks



(a) Common failure mode where the path wanders away from the breast-air boundary, towards another edge in the image with low intensity.

(b) Failure to segment the right breast.

Figure 3.1.2: Failure modes of the Hayton breast-air segmentation algorithm.

simultaneously using a graph-cut method. In this method, the inherently three dimensional problem is solved as a three dimensional problem, as opposed to reducing the problem into two dimensional stages.

Graph cuts has become a popular method for low level processing functions such as image segmentation and restoration. Graph cuts is an energy minimization technique based on a combinatorial optimization. Graph cuts provides a solution that is globally optimal, and through it's tunable energy function it can be used widely.

In the graph cut approach, images are described as non-oriented graphs, where the nodes on the graph correspond to voxels and the arcs correspond to the adjacency relations between voxels (or indeed supervoxels). Two additional nodes are added to the graph, which correspond to the source s and the sink t. The method is initiated by the interactive or automated identification of one or more nodes on the graph which represent the object and the background. These nodes are called seeds, and serve as hard constraints in the binary segmentation of the image. Soft constraints are also added, which represent regional or boundary information. The s and t terminals are hard-linked to the seeds which correspond to the the object and background respectively. Therefore, three types of arcs are defined in the graph: $n-links$, $s-links$ and $t-links$: $n-links$ connect neighbouring pairs of nodes whose costs are defined by the regional or boundary information, while $s-links$ and $t-links$ link nodes to the terminals and have costs derived from their similarity to the terminal in question. An example graph is shown in Figure 3.2.1.

A $s-t$ cut in graph G is a set of arcs whose removal partitions the graph into two disjoint subsets S and T such that no path can be created between s and t. The cost associated with this cut is the total cost of all the arcs in the cut, and the minimum $s-t$ cut is therefore a cut whose cost is minimal. The minimum

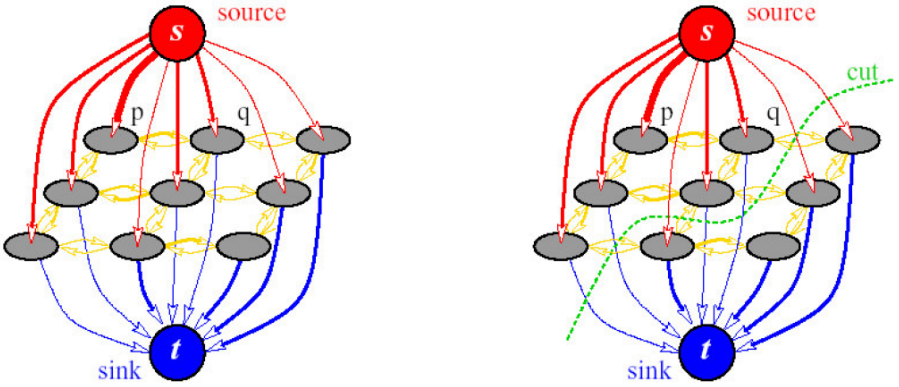


Figure 3.2.1: Graph Cuts Schematic
Left: Graph shown with $n-links$ and $t-links$, as well as source and sink nodes. Right: Binary partition of the graph. Figure adapted from [38].

$s-t$ cut problem has a dual problem, the maximum flow problem, and these are classic combinatorial problems that can be solved by number of different algorithms (e.g. Ford & Fulkersen, Edmonds-Karp [39]). The maximum flow problem can be visualized by imagining the source to be a water source like a tap, the arcs of the graph as pipes whose cost corresponds to their capacity, and the sink to be a drain. Then, the problem is reimagined as the maximum amount of water that can flow from the source through the network of pipes to the sink. An overview of the algorithm is presented in Algorithm 3.2.

The largest bottlenecks in applying graph cuts are computational. The algorithm has a worst case complexity of $O(|E| \cdot |V|^2)$ where $|E|$ is the number of edges and $|V|$ is the number of nodes. Using pre-segmented regions as opposed to voxels can reduce the number of nodes by several orders of magnitude which therefore significantly speeds up the processing. Similarly, the memory requirements for graph cuts can be very large and prohibitive as the size of the volume increases. The popular graph cuts implementations of [40, 41, 42] requires $40V + 32E$ bytes to store the graph on a 64 bit machine [41, 34].

For example, using full 26-connectivity on a typical registered breast MRI volume, this yields a memory footprint of approximately 30.85 [6] gigabytes excluding any additional overhead. Performing graph cuts on a graph defined on supervoxels can reduce this memory footprint by several orders of magnitude. In a similar approach, Stawiaski et al. [43] perform graph cuts on top of an oversegmentation generated by the well known watershed algorithm. A region adjacency graph[7] is defined on the watershed regions on the image. Stawiaski et al. note the high memory requirements for the method of [41] when attempting to segment a thoracic CT volume, which in fact meant that they could not compute the segmentation for use as a comparison to their own work. Using the SLIC segmentation method as preprocessing, the same typical, registered breast MRI volume was able to be reduced to a much more manageable 0.031

[6] Calculated using 40V+32E, on a volume with size $512 \cdot 512 \cdot 150$

[7] A region adjacency graph defines how regions (watershed regions, mean shift clusters, supervoxels) which correspond to nodes in the graph are connected to other neighbouring regions.

Algorithm 3.2 Graph Cut Segmentation

Input: Image (or volume)

Create an arc-weighted, directed graph corresponding to size and dimensionality of image to be segmented

Identify object and background seeds

Create two special graph nodes: source s and sink t, and connect all identified seeds with either the source or sink based on their respective background or object label

Assign Arc costs for $n-links$, $s-links$, and $t-links$

Compute the optimum cut using one of the maximum flow graph optimization algorithms

Visualize the cut by relabelling the original image to correspond with the new background and object labels from the cut

Output: Binary labelled image (or volume)

gigabytes [8].

The proposed method is based on that of by Stawiaski et al., except that instead of building a region adjacency graph on 2D watershed regions, here the region adjacency graph is defined on supervoxel labels and the entire operation is performed in 3D. Another difference is that while Stawiaski et al. used an interactive process to create hard-constraints, here it is automatic. Overall, the process can be seen as a region merging method.

First, the capacities (arc costs) need to be defined for the $n-links$, $s-links$, and $t-links$. The $s-links$ and $t-links$ reflect the similarity of the node into one of the two sets (S,T), corresponding to the background or object. Stawiaski et al. define:

$$c(s,i) = \frac{d^O}{d^O+d^B} \quad \forall i \epsilon \in I \tag{3.2.1}$$

and

$$c(i,t) = \frac{d^B}{d^O+d^B} \quad \forall i \epsilon \in I \tag{3.2.2}$$

where $d^O = |m_O - m(i)|$, $d^B = |m_B - m(i)|$ and I is the set of all voxels in the volume. The values m_O and m_B are the average intensity levels of the regions marked respectively as object and background, and $m(i)$ is the average intensity in region (supervoxel) i. The values m_O and m_B could be defined though user input (as in Stawiaski et al) or through an automatic method.

In this case, given that air produces a near-zero MRI intensity, establishing an estimate for the background intensity level is simple. For example, a threshold percentile could identify the lowest 1% of supervoxel intensities and assume

[8] Using the same calculation method as before, and supervoxels generated for a complexity reduction of approximately 710x. This estimate does not include ancillary variables needed during the MATLAB computations and simply reflects the reduced number of nodes and arcs in the graph.

this to be representative of air. Alternatively, from a histogram of supervoxel intensities, a bimodal distribution would be expected, where the first mode corresponds to supervoxels who have an intensity near to that of air and the second mode corresponding to the most common breast tissue signal. The proposed method uses the latter approach, setting the average supervoxel intensity for air, m_B to the first mode of the distribution. The estimated intensity average for breast tissue, m_O , is estimated by using the second mode of the distribution.

Next, to create the capacities for the $s-$ and $t-links$ a variation on the proposed similarity metric of equations 3.2.1 and 3.2.2 is used:

$$c(s,i) = \alpha \cdot \exp(-\frac{d^O}{d^O+d^B}) \quad \forall i\epsilon \in I \tag{3.2.3}$$

and similarly,

$$c(i,t) = \alpha \cdot \exp(-\frac{d^B}{d^O+d^B}) \quad \forall i\epsilon \in I. \tag{3.2.4}$$

An exponential function was found to produce better results as opposed to the simple ratio of Stawiaski et al. α is a scalar tuning parameter, which can be used to scale the relative weights between the capacities of $s-$ and $t-links$ versus $n-links$.

Using the estimates in intensity for breast tissue and air, seed points which act as hard constraints are created. These seed points are always included in the set which they represent (i.e. air seeds are always included in the air segment of the image). To do this, the capacity of the arc should be defined in a way where it is never advantageous to cut it. Therefore, they are assigned an infinite capacity which provides this hard-constraint. The seed points also ¨disconnected¨ from the opposite label by assigning them a capacity of zero. Therefore:

$$\begin{cases} c(i,s) = 0 & \forall i \in I - \{O \cup B\} \\ c(i,t) = 0 & \forall i \in I - \{O \cup B\} \\ c(i,s) = \infty & \forall i \in \{O\} \\ c(i,t) = \infty & \forall i \in \{B\} \end{cases}, \tag{3.2.5}$$

where O is the set of breast (object) seeds, and B is the set of air (background) seeds. The air seeds are determined by selecting all nodes which have an intensity less than the m_B estimate. The breast seeds are chosen as the nodes that have intensity values in the range $m_O \leq m(i) \leq 2 \cdot m_O$.

Finally, the $n-links$ are assigned capacities. These capacities should reflect the dissimilarity between regions [43, 38, 41]. Many methods are available to measure the dissimilarity between the regions, including: the average intensity, the histogram of intensity values, the cumulative histogram, or information from the region's boundary. In this work, a metric based off of the region's average intensity is used:

$$c(i,j) = \beta \cdot \exp(-|\frac{I_i - I_j}{(\max(I) - \min(I))}|), \tag{3.2.6}$$

where I_i is the intensity of the region i, and β is again a scalar tuning parameter similar to α. The maximum intensity difference between any two supervoxels in the image is used to normalize the dissimilarity metric.

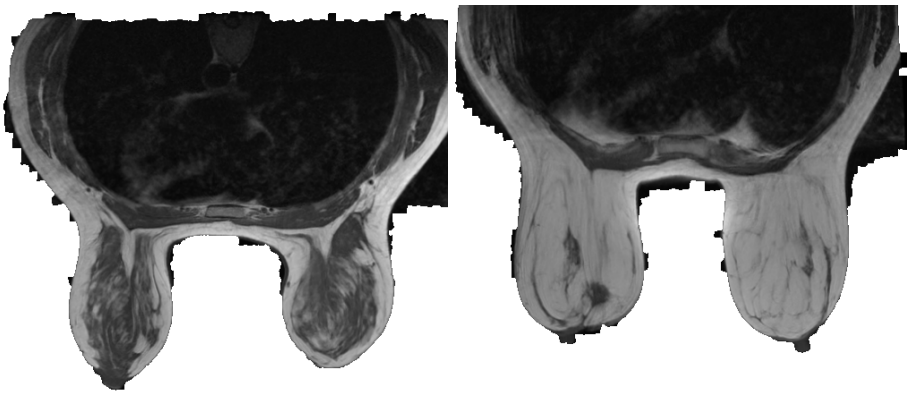


Figure 3.3.1: Sample segmentations from the proposed supervoxel-based algorithm on T1-weighted images.

3.3 Example Segmentations from Proposed Method

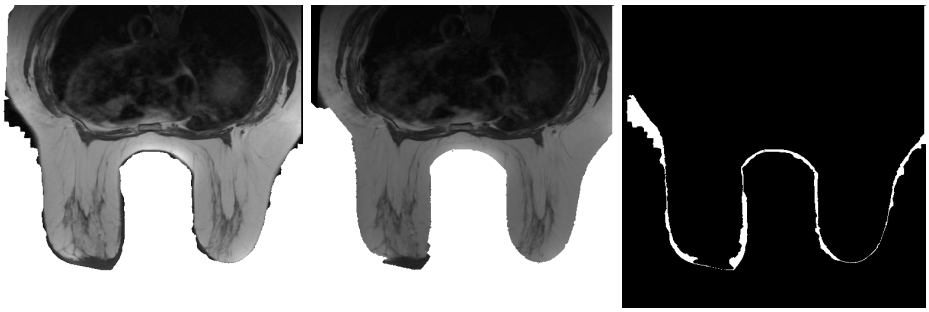
In this section the resulting segmentations from the supervoxel graph-cuts approach are shown. Experimentally, α and β were determined as 4 and 1 respectively after testing on a range of different volumes. They were selected as they produced stable output segmentations.

Figure 3.3.1 shows sample output using the proposed method. Here, the areas representing air have clearly been removed, and the boundary of the breast has been maintained. The nipples have also been preserved in the some of the segmentations. Figure 3.3.2 shows a comparison between the Hayton algorithm and the proposed method for a typical slice. In Figure 3.3.2b an area of low intensity exists near the nipple of the left breast, and here the Hayton segmentation has produced a jagged line segmentation. This type of failure mode is common with the Hayton algorithm, because it uses a dynamic programming technique and must search pixel-by-pixel it can be led astray. In the same section, the proposed method did not produce such an unnatural contour.

However, the proposed method is not perfect, and similar jagged outcroppings on the boundary exist as the supervoxel on the boundary is included in the cut. These occurrences are likely because the mean intensity inside that supervoxel is higher than what would be expected for air, and so it is kept in the cut. This could be due to partial-volume effects which would increase the intensity in the boundary region.

As the proposed algorithm relies on previously generated supervoxels, the total runtime is similar to Hayton's method [9]. If the generation of the supervoxels was considered a separate stage, and not included in the total runtime, the proposed method would be approximately 50x faster. The proposed method has

[9] (approximately 600s for a volume of size 512x512x150, on a Mac OSX, 2.5 GHz Intel Core i5 system)



(a) Proposed Algorithm (b) Hayton's algorithm (c) Subtraction image of the two binary masks

Figure 3.3.2: Comparison between Hayton Segmentation and Proposed Algorithm for T1-weighted images.

a runtime of approximately 5-10 seconds if the initial segmentation is excluded. This comparison could be fair if the large-scale segmentation used in this process was also used for subsequent tasks. For example, this high-level segmentation could be used in an algorithm to identify suspicious areas which are then segmented again at a finer level for further analysis. This hierarchical strategy may become more important in the future if the size of the volumes increases, as it will be come more difficult to manage the large datasets. Furthermore, as the size of the volume increases, the Hayton algorithm becomes increasingly expensive to compute as the two cost maps are larger and the neighbour-search stage is forced to search through more nodes as the path lengthens. From this perspective, the ability to reduce the complexity of the volume is useful and supervoxels may prove more valuable in the future.

The proposed method's performance may be improved by using other metrics for dissimilarity between regions. For example, histogram based methods could be used, where are an approximate measurement of histogram overlap or similarity would be used. Functions based on the boundaries of the supervoxels could also be implemented as for example, two supervoxels are more likely to belong to the same set if their shared boundary is very similar in intensity.

The proposed method is general and can be applied to various other problems outside of breast-air segmentation. A similar application of supervoxels and graph cuts has been reported in brain MR analysis [33], where the supervoxels are later merged downwards to represent three classes of tissue in the brain (White Matter, Gray Matter, CSF). When applying the proposed algorithm to a T1-weighted brain volume to create a mask of the skull, Figure 3.3.3 was obtained in approximately 45 seconds including the initial SLIC segmentation. Graph-cuts is extensible to $n - class$ problems, and it would be interesting to apply this in breast MRI to segment the breast-air boundary and chest-cavity simultaneously in 3D.

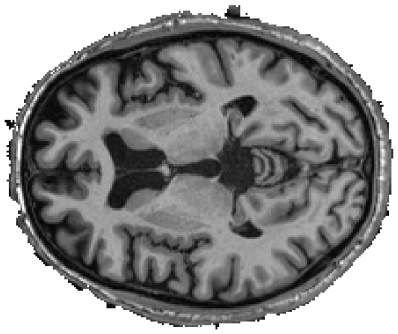


Figure 3.3.3: Example segmentation of the skull-air boundary using the proposed method.
$k = 20000$ and $m = 20$. The T1-weighted brain volume was of size $240 \times 240 \times 48$.

Chapter 4

Multimodal Segmentation of Breast Lesions

This chapter, and the two chapters that follow address the third objective of this thesis: to develop and evaluate supervoxel-based features for classifying breast lesions as benign or malignant. In this chapter a method is presented for partitioning the voxels inside a predefined volume-of-interest (corresponding to a suspicious lesion) into supervoxels based on multimodal MRI values. In the next chapter this partitioning is used to derive features for quantitatively characterizing the lesion. This chapter is organized as follows. First, the multimodal SLIC implementation used to partition a lesion into supervoxels is described. Next the method used to tune the parameters of the algorithm for use on the data outlined in Section 1.4 is described.

4.1 Multimodal Implementation of SLIC

In this section the multimodal implementation of the SLIC algorithm used for partitioning a clinician-defined volume-of-interest (VOI) into supervoxels is described. These VOIs are small binary masks that correspond to voxels that are within a suspicious region of tissue. A typical VOI is shown on a T1-weighted image in Figure 4.1.1.

The proposed algorithm takes as input the full set of DCE-MRI volumes acquired in the axial plane. Let the vector of signal intensities at each voxel be denoted:

$$I_{x,y,z} = \begin{bmatrix} I_{t_0} & I_{t_1} & I_{t_2} & I_{t_3} \end{bmatrix}^T, \tag{4.1.1}$$

where I_{t_0} is the pre-contrast volume, and I_{t_1} and subsequent time-points are post-contrast. Using this intensity vector, the intensity distance of voxel i to cluster centre j becomes:

$$d_I = \sqrt{(I_{t_0 i} - I_{t_0 j})^2 + (I_{t_1 i} - I_{t_1 j})^2 + (I_{t_2 i} - I_{t_2 j})^2 + (I_{t_3 i} - I_{t_3 j})^2}. \tag{4.1.2}$$

Each post-contrast volume is defined using the procedure defined by Stoutjesijk et al[16]. First the relative enhancement RE of voxel v at time t , is computed

(a) T1-Weighted image with VOI overlaid (b) Delineation of Lesion

Figure 4.1.1: Typical lesion presentation with VOI overlays from both T1-weighted and first post-contrast images.

viz:

$$RE(v,t) = \frac{SI(v,t) - SI(v,t_0)}{SI(v,t_0)} \cdot 100\%, \tag{4.1.3}$$

where $SI(v,t)$ denotes the signal intensity of voxel v at time t. Then normalizing:

$$RE'(v,t) = \frac{RE(v,t) - \mu}{\sigma}, \tag{4.1.4}$$

where μ and σ are the mean and standard deviation of $RE(v,t)$ over all voxels.

Similar to [19] three different cluster sizes are defined for analysis: small, medium, and large. These cluster sizes are created using different bandwidths in the mean shift algorithm. Here, using SLIC's spacing parameter S, the average number of voxels in each supervoxel is estimated as S^3[4].

Next, the compactness parameter m should be determined. The compactness parameter can also affect how many voxels are assigned to each supervoxel, although it is expected that as it increases the supervoxels will become more compact and regularly shaped as the distance metric tends to favour the spatial distance.

SLIC segmentation for the DCE-MRI volume set is summarized in Algorithm 4.1.

Algorithm 4.1 Multimodal Simple Linear Iterative Clustering within VOIs

Input: DCE-MRI Volumes (1 Pre-contrast, 3 Post-contrast), desired supervoxel size (small, medium,large), compactness parameter m

Initialize cluster centres $C_k = \begin{bmatrix} I_{t_0} & I_{t_1} & I_{t_2} & I_{t_3} & x & y & z \end{bmatrix}^T$ by sampling at regular grid intervals S

Perturb cluster centres to lowest gradient position in $3 \times 3 \times 3$ region from pre-contrast t_0 volume

Set labels $l(i) = -1$ for each voxel i

Set distance $d(i) = \infty$ for each voxel i

Repeat

for each cluster centre C_k
for each voxel i in $2S \times 2S \times 2S$ region around C_k
if voxel is within VOI
Calculate distance D between C_k and i
if $D < d(i)$ then
set $d(i) = D$
set $l(i) = k$
end
end
end
end
Compute new cluster centres
Compute residual error

until Error $\leq$ Threshold or Number of Iterations reached

Enforce Connectivity

Output Labelled Volume where voxels outside of the VOI are label #1

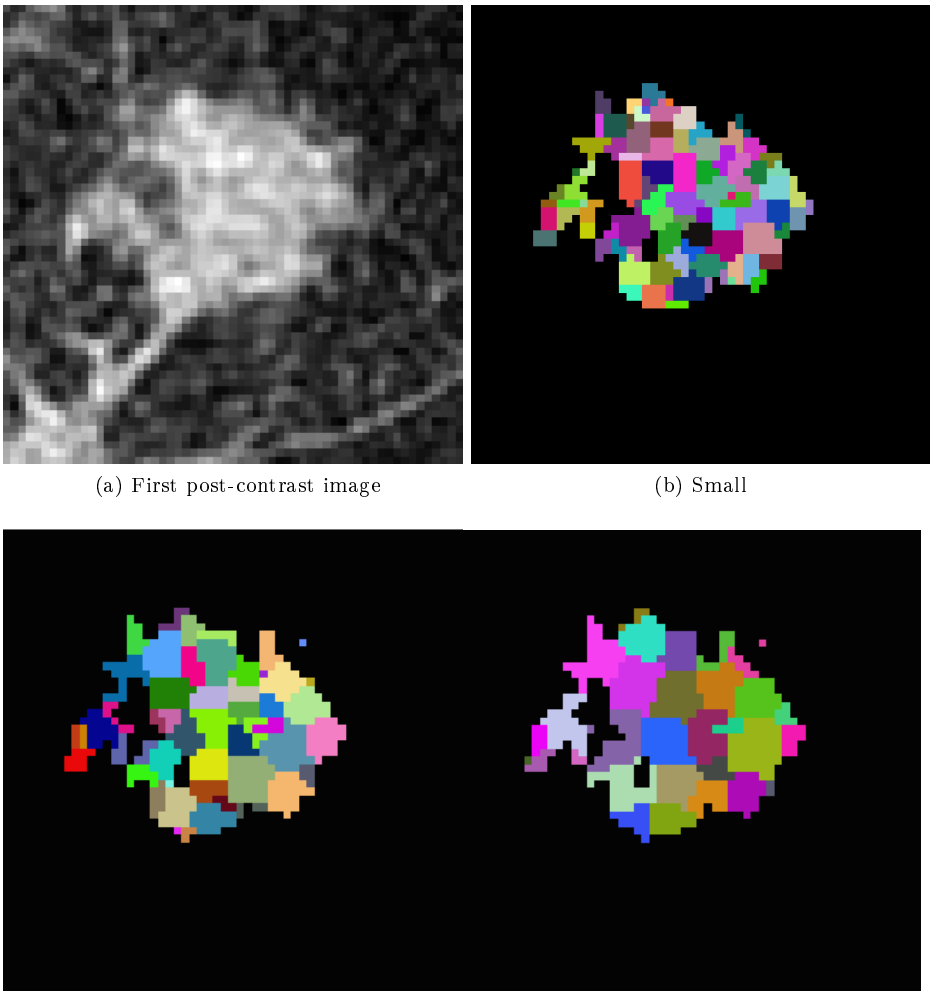
4.2 Application to the clinical data

This section presents the specific parameter values tuned to the clinical data used in this thesis.

Spacings of 3, 5, and 7 were used for the three segmentation sizes, which correspond to on average 27, 125 and 343 voxels per supervoxel respectively [1]. A minimum supervoxel size was defined as 5 voxels, and any resulting segments which were less than this threshold were merged to the closest neighbour of

[1] Segmentation sizes were based off of the work of [19] who used similar size definitions while segmenting using Mean Shift.

similar intensity. Figure 4.2.1 shows a sample segmentation for each size and the first post-contrast image for comparison.



(a) First post-contrast image (b) Small

(c) Medium (d) Large

Figure 4.2.1: VOI segmentations at various supervoxel sizes. Segmentations were performed using $m = 5$ at $S = 3, 5, 7$ (small, medium, large). Note: each supervoxel is assigned a random colour, and the background is black. This VOI was clustered into 567, 212, 136 supervoxels for small, medium and large sizes respectively.

Next, the compactness parameter m was determined. To do this, a visual inspection of several segmentations was performed. Segmentations using $m = 1, 2, 5$, and 10% were compared using the same spacing parameter ($S = 3$). Sample segmentations performed for this inspection can be seen in Figure 4.2.2. Here, it appears that $m = 5$ is the most visually pleasing with respect to boundary adherence compared to the normalized post-contrast image. This value also seems to yield the best balance between number of supervoxels, their shape, and

their adherence to boundaries in the comparison image. Therefore, $m = 5$ was selected for further use.

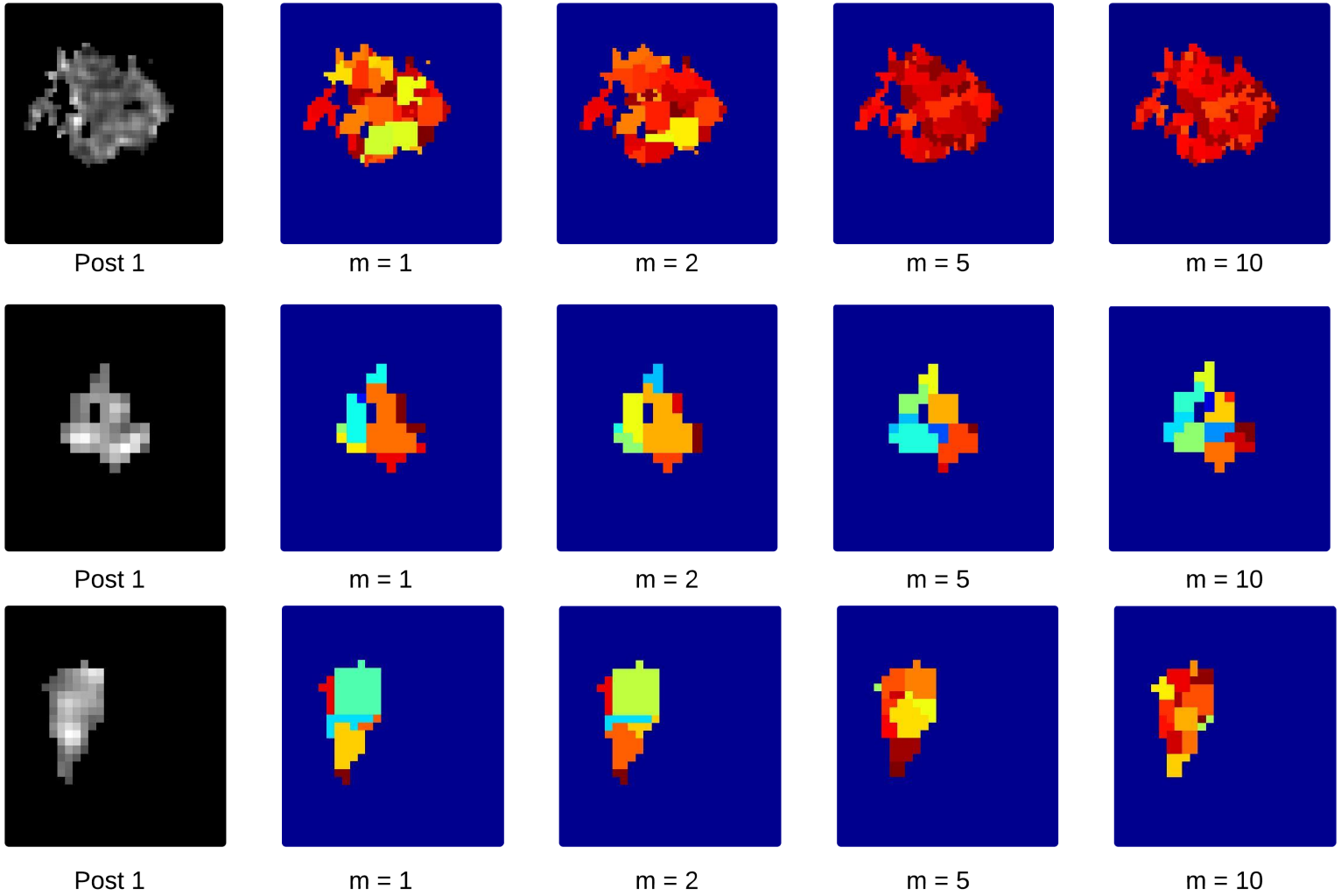


Figure 4.2.2: VOI Segmentations at various compactness parameter settings.
Segmentations were performed using $S = 3$. Note: each colour represents a supervoxel. After visual review, compactness parameter $m = 5\%$ was chosen.

Chapter 5

Supervoxel-based features for characterizing breast lesions

This chapter builds on the previous to address the third objective of this thesis: to develop and evaluate supervoxel-based features for classfiying breast lesions as benign or malignant. Specifically, this chapter presents several features derived from the partitioning of a lesion into supervoxel clusters using the algorithm of chapter 4.

The features that are described in this chapter stem from cluster-based features that have been defined for other partitioning schemes in the literature, such as mean-shift [19].

The chapter is organized as follows. Section 5.1 discusses the DCE-MRI based features that were used, and Section 5.2 discusses the DW-MRI based features that were used.

5.1 Supervoxel-based DCE-MRI Features

Dynamic contrast enhanced imaging studies are common in breast MRI exams. It has been well established that malignant lesions release angiogenic factors that induce growth of new and preexisting vessels [44]. However, these vessels exhibit a pathological wall structure, with leaky endothelial linings. The affect of the angiogenic activity is twofold: there is an increased density of vessels which leads to an increased flow of contrast material, and an increased vessel permeability which causes the contrast material to leak out into the surrounding tissue. Fast uptake of contrast material, as well as fast washout are both often anecdotally described as being signs of malignancy[18].

To describe the uptake, and subsequent washout of contrast material various models have been proposed [6, 35, 22, 24, 23]. These models are shown in Table 5.1. These models were not fitted voxel-wise, but to clusters of voxels defined by the SLIC segmentation on the relative enhancement images. They are also calculated for each segmentation cluster size (small, medium, large).

After fitting the data to the models, the features used in classification are respectively:

- the Linear slope parameters β_1, and β_2,

Table 5.1: DCE-MRI parametric models of enhancement.

Note: $t_j$is defined as the time (in seconds) since contrast material injection for the jth DCE-MRI stack, and $t_0 = 0$. † - From Ricker Model

Parametric Model	Equation	Initialization for non-linear least squares fitting	Bounds
Linear Slope model [24]	$f(t) = \begin{cases} \beta_1 t & t \leq \alpha \\ \beta_1\alpha + \beta_2(t-\alpha) & t > \alpha \end{cases}$	$\beta_1 = \frac{R_{C_i}(1)-R_{C_i}(0)}{t_1-t_0}$ $\beta_2 = \frac{R_{C_i}(M)-R_{C_i}(1)}{t_M-t_1}$ $\alpha = t_1$	$-\infty < \beta_1 < \infty$ $-\infty < \beta_2 < \infty$ $-\infty < \alpha < \infty$
Hayton model [35]	$f(t) = \frac{A}{a-b}(e^{-bt} - e^{-at})$	$a = \frac{0.2}{60}$ $b = \frac{0.1}{60}$ $A = \frac{R_{C_i}(M)-R_{C_i}(1)}{e^{-bt}(1-e^{-at_2})}$	$0 \leq \alpha \leq 0.1$ $-5 \cdot 10^{-4} \leq b \leq 0.2$ $-0.01 \leq A \leq 0.01$
Ricker model [23]	$f(t) = \alpha t e^{-\beta t}$	Least Squares	$0 \leq \alpha \leq 1$ $-0.01 \leq \beta \leq 0.05$
Simplified gamma variate model [22]	$f(t) = At^{\alpha}e^{-t/\beta}$	$A = \alpha$† $\alpha = 1$ $\beta = 1/\beta$†	$0 \leq A < \infty$ $0.3 \leq \alpha < \infty$ $50 \leq \beta < \infty$
Agliozzo model [6]	$f(t) = Ate^{-t^D}$	$A = \alpha$† $D = 0.1$	$-\infty < A < \infty$ $-\infty < D < \infty$

- the Hayton model parameters A, α, and b,
- the Ricker model parameters α, β,
- the Simplified Gamma Variate model parameters A, α, and β,
- and the Agliozzo model parameters A and D

In the case of the linear slope model parameter β_1 and the simplified gamma variate model parameter α the 10^{th} percentile is extracted. For all other model parameters, the 90^{th} percentile is extracted. This is done to identify the most malignant-appearing portion in the lesion, while reducing the effect of outliers. In total, 36 DCE-MRI derived features are created for each VOI, where the 12 features listed above are calculated for each cluster size (small, medium, large). They are assigned the labels $D1 - D36$ where they are ordered as shown in Table 5.2.

5.2 Supervoxel-based DW-MRI Features

It has been shown that malignant lesions have a lower mean apparent diffusion coefficient (ADC) than benign lesions[45, 46]. However, the calculation of ADC is highly sensitive to noise and geometric distortions when performed on a voxel-by-voxel basis due to eddy currents induced by strong diffusion gradients. This is the motivation for creating features on a cluster basis. To create these cluster based features first the voxel-wise ADC must be calculated as follows:

$$f_{ADC}(x) = -\frac{1}{b} log(\frac{f_{DW_b}(x)}{f_{DW_0}(x)}), \tag{5.2.1}$$

Table 5.2: DCE-MRI Features

Note: each feature triplet is ordered by cluster size, where for example: $D1$ corresponds to Least Squares β_1 for Small Clusters, $D2$ for medium clusters, and $D3$ for large clusters.

Label	Feature
$D1-3$	Linear Slope β_1
$D4-6$	Linear Slope β_2
$D7-9$	Hayton A
$D10-12$	Hayton α
$D12-15$	Hayton b
$D16-18$	Ricker α
$D19-21$	Ricker β
$D22-24$	Simplified Gamma A
$D25-27$	Simplified Gamma α
$D28-30$	Simplified Gamma β
$D31-33$	Agliozzo A
$D34-36$	Agliozzo D

where x is the list of voxels inside the VOI, $f_{DW_0}(x)$, and $f_{DW_b}(x)$ are the diffusion weighted stack for b-value zero and b respectively. To create features using the less noisy cluster-based data, let $C_i \subseteq V$ represent the ith cluster. The mean of the values in $f_{DW_b}(x)$ within the cluster is given by:

$$\mu_{C_i, f_{DW_b}} = \frac{1}{|C_i|} \sum_{x \in C_i} f_{DW_b}(x). \tag{5.2.2}$$

The value $\mu_{C_i f_{DW_0}}$ is similarly defined. The ADC, α, for cluster C_i is calculated as follows:

$$\alpha_{C_i} = -\frac{1}{b} log(\frac{\mu_{C_i, f_{DW_b}}}{\mu_{C_i, f_{DW_0}}}). \tag{5.2.3}$$

Now that the ADC has been computed for each cluster, the following are calculated for each cluster size: minimum cluster ADC, the 5th, 10th, 25th, 50th cluster ADC percentiles, the maximum cluster ADC, the mean cluster ADC, and the mode cluster ADC.The mean ADC in the cluster of maximum initial mean enhancement is also computed using cluster parameter information from the DCE-MRI feature list. First the initial contrast enhancement is calculated for each cluster using:

$$\beta_{C_i} = \frac{\mu_{C_i, f_{DCE_1}} - \mu_{C_i, f_{DCE_0}}}{\mu_{C_i, f_{DCE_0}}}. \tag{5.2.4}$$

Then the cluster with maximum initial contrast enhancement is $C_{i_{max}}$,where:

$$i_{max} = \max(\beta_{C_i}) \quad i \in \{1, ..N\} \tag{5.2.5}$$

and this is performed identically for small, medium, and large cluster sizes.

The mean ADC in the cluster of maximum mean DW-MRI intensity is also used as a feature for each cluster size. The cluster with the highest DW-MRI intensity is $C_{k_{max}}$where:

$$k_{max} = \max(\mu_{C_k, f_{DW_0}}) \quad k \in \{1, ..N\}. \tag{5.2.6}$$

Finally, the whole lesion mean ADC is also used as a feature. This is the only feature that is not cluster based which is used. This is trivially α_V where V is the whole VOI.

In total, 31 DW-based features are created for each VOI, where the features listed above are calculated for each cluster size (small, medium, large). They are assigned the labels $W1 - W31$ where they are ordered as shown in Table 5.3.

Table 5.3: DW-MRI Features

Label	Feature
$W1 - 8$	min, 5th, 10th, 25th, 50th %, max, mean, mode ADC for small clusters
$W9 - 16$	min, 5th, 10th, 25th, 50th %, max, mean, mode ADC for medium clusters
$W17 - 24$	min, 5th, 10th, 25th, 50th %, max, mean, mode ADC for large clusters
$W25$	Whole lesion mean
$W26, W28, W30$	ADC in cluster of maximum initial mean enhancement (S,M,L)
$W27, W29, W31$	ADC in cluster of maximum mean DW intensity (S,M,L)

Chapter 6

Lesion Classification

This chapter together with the two preceding chapters addresses the third objective of this thesis: to develop and evaluate supervoxel-based features for classfiying breast lesions as benign or malignant. Specifically, this chapter describes several classification experiments to determine both the discriminatory performance of the proposed features of chapter 5, and the most important subset of features. This chapter is organized as follows. Section 6.1 describes the classification algorithm used, Section 6.2 describes the classification experiments, and Section 6.3 presents the results of the classification experiments.

6.1 Random Forest

The classification algorithm chosen to evaluate the performance of the proposed supervoxel-based features and to identify the most significant subset of these is the Random Forest (RF).

RF is a versatile classification algorithm for the analysis of large data sets, which uses an ensemble of classification trees. Each classification tree is built using a bootstrap sample of the data set, and at each split in the tree the candidate set of variables is a random subset of the larger group of variables. Therefore, RF uses both bootstrap aggregating (bagging), which is known to be a successful approach for aggregating weak or unstable learners, and random variable selection for tree building. RF was created by Leo Breiman, in 2001[47]. The algorithm yields an ensemble that can achieve both low bias and low variance (due the the averaging process over a large number of trees who have low bias themselves, but high variance) [48, 49, 47]. RF has shown excellent performance during classification applications and can have performance comparable to support vector machines[48].

RF has several very useful characteristics:

1. Can be used effectively when there are many more predictive variables (or features) than observations. This reduces the effect of the curse of dimensionality.
2. Can be used for binary classification, or multi-class problems.
3. Has good predictive performance even when most predictive variables are noise.

4. Does not overfit to the training data and is relatively robust to outliers and noise[47, 48, 49].

5. Can handle a mixture of categorical and continuous predictors.

6. Incorporates interactions among predictor variables.

7. The output is invariant to monotone transformations of predictor variables.

8. Returns measures of predictor variable importance.

9. There is little need to fine-tune parameters to yield best performance. There is only one important parameter to tune, *mtry*, which is the number of variables tried at each split in the tree. However, it has been reported that the default setting is often a good choice, and that the tuning effect is often minimal[47, 49, 50]. Only two other parameters need to be decided on: the number of trees to grow in the forest, *ntree* and the minimum size of the terminal nodes *nodesize*.

10. High quality and free implementations exist, and the algorithm can be easily parallelized.

The RF algorithm is a collection of trees (a forest) with variations in structure, generated via two modifications to the well known deterministic tree growing algorithm [47, 51].

First, the best split at each tree node is selected among a random subset of predictor features. Secondly, the training set used to grow each tree is a bootstrap sample of the total number of observations N, and is drawn with replacement from the original set of samples. Most often, this random sample is approximately 63% of the data[49]. Some observations could therefore be represented multiple times in the training set, while others may not be included. The observations which are not included are described as being "out-of-bag"(OOB). Those observations which are OOB are used to estimate the error in prediction.

Each of the trees in the random forest are grown to their full possible extent. Different predictor variables (features) are used at each split in each of the trees. These predictor variables are chosen at random from the full set of variables, and are then evaluated for their ability to split the data. The variable which results in the largest decrease in impurity (or largest information gain) is chosen to separate the samples at that node, resulting in two distinct child nodes. This process continues until the final nodes contain only samples belonging to the same class, or contain a specified minimum number of samples. Usually a classification tree is grown until the terminal nodes are pure, even if that means that the final node contains only a single sample. This process repeats until the user specified number of trees is grown, and is depicted in Figure 6.1.1.

To predict the class of an observation, the observation is assigned to an end-node in the tree (a leaf) based on its predictor values. The class of the majority of training set observations in the leaf is selected as the classification predication for this specific observation. Using the whole forest of classification trees, each tree is assigned one unweighted vote for each OOB observation, and for each observation the class receiving the most votes is the predicted class. Ties, whose probability of occurrence is quite low, are resolved by selecting the class with the lowest label. This process is shown in Figure 6.1.2.

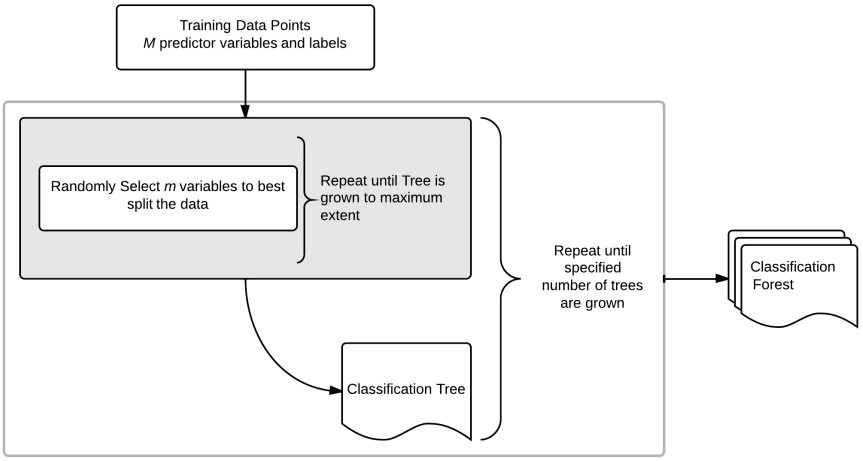


Figure 6.1.1: Creation of a Random Forest of classification trees.

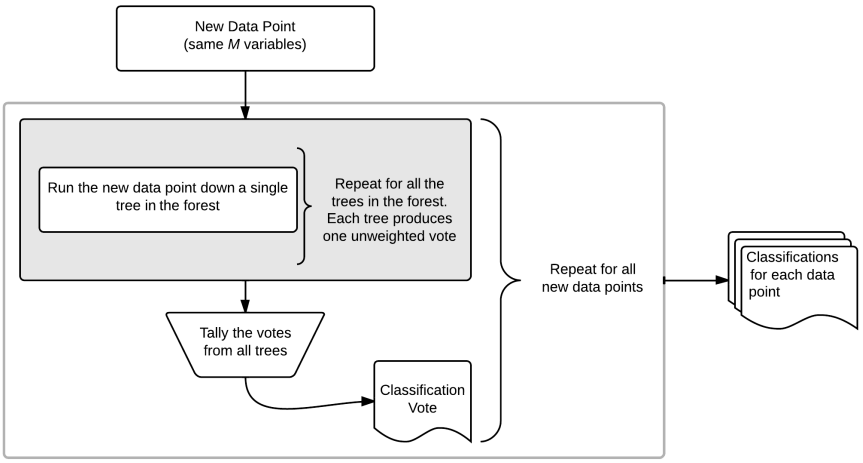


Figure 6.1.2: Classification of a new data point using a Random Forest of classification trees.

General cross-validation[1] procedures are not needed when predicting the classification performance of RF. A cross-validation scheme is already built in as each tree in the forest has its own training (bootstrap) and test (OOB) datasets[49, 47]. The random selection of candidate predictor variables during the splitting procedure at each node ensures a low correlation between the trees and prevents over-training[2] of the RF to the data. An outline of the algorithm is provided in Algorithm 6.1.

Algorithm 6.1 Random Forests

For each tree to *ntrees*

Draw a bootstrap sample of size N from the data, with replacement.

Grow a random-forest tree to the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree until minimum node size *nodesize* is reached, or the node is pure

Select m predictor variables at random from the full set

Pick the best predictor variable and split point among the m variables

Split the node into two child nodes

Output the forest of trees

Classify sample x as the class of the majority vote from all trees in the forest

---

Data sets with many weak inputs are especially common in classification problems inside the life sciences. For example, thousands of genes (predictor variables) being used for only a few dozens of samples. In these scenarios, RF performs very well, and since it can output measures of variable importance, that information could be used to filter out non-informative variables or identify biomarkers for disease. For this reason, a randomly generated feature is created and added to the list of predictor variables, which has no diagnostic importance. This can be thought of as a threshold for the importance scores, as a feature that is important should be at least more important than something which inherently has no diagnostic ability.

6.2 Classification Experiments

After performing the SLIC segmentation of the images for each supervoxel size, the features described in Tables 5.2 and 5.3 are calculated. A randomly generated feature was created using MATLAB's built in normally distributed pseudorandom number generator. The features were compiled into a spreadsheet

[1] Cross-Validation is a set of techniques used to assess how a model will generalize to an independent data set. Generally, these techniques involve partitioning the data-set into smaller subsets for testing and validation.

[2] Over-Training, or Over-Fitting, describes when the model has poor predictive performance because it describes random error or noise in the data, instead of the underlying relationship. Over-trained models often perform well using the data they were trained with, but poorly with another independent set.

Algorithm 6.2 Classification Experiment Script Pseudocode

Import Features spreadsheet (comma separated values)

Create Labels for DCE-MRI, DW-MRI features, random feature, and class label

Define Number of experiment repetitions, *mtry*, *ntree*

Repeat

> **Create** Random Forest without regularization using `RRF` package, specified *mtry*, *ntree*, and features
>
> **Save** RF importance scores
>
> **Create** ROC using `proc` package
>
> **Save** AUC

Calculate Mean AUC±SE, Mean importance scores per feature

Plot ROC

sorted by subject number, and supervoxel size. Finally, each lesion's known classification label was appended to the spreadsheet. The features spreadsheet is read by a script creating the RF in the R software package [27]. This script creates and saves the receiver-operating characteristic (ROC) curves in both the raw, empirical manner and a smoothed version (default, binormal smoothing is used). Each classification experiment is run 100 times, using the specified *ntree* and *mtry* parameters, and the mean area under the curve (AUC) is reported with the standard error (SE). An outline of the R script is shown in Algorithm 6.2. A more detailed description of the script can be found in [19].

The standard test parameters of $ntree = 2000$ and $mtry = 3$ were set by [19] using similar data.

Five classification experiments were performed:

1. Classification using all features, using $ntree = 2000$, and $mtry = 3$;
2. Classification as in 1) but for various values of $mtry$:$\{2, 3, 4, 5, 10, 20\}$;
3. Classification as in 1) but for various values of $ntree$:$\{500, 2000, 3000, 4000, 10000\}$;
4. Classification as in 1) but using only the DCE-MRI features; and
5. Classification as in 1) but using only the DW-MRI features.

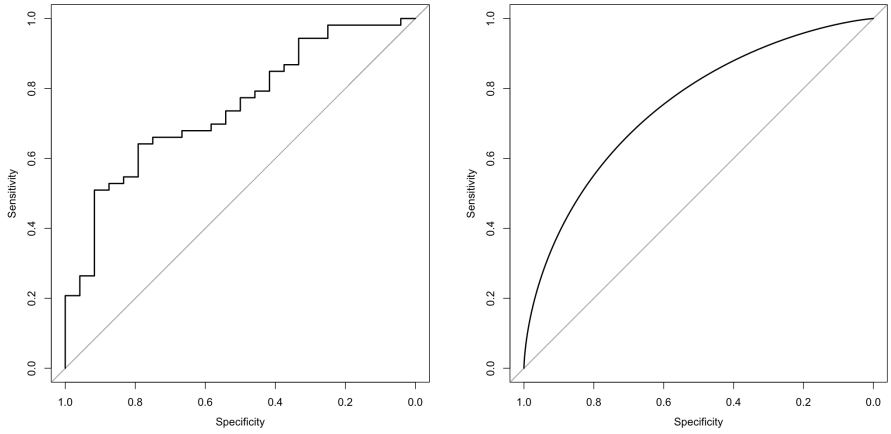
6.3 Results and Discussion

The results from the classification experiments are presented in terms of the area under the curve (AUC) of the receiver-operating characteristic curve (ROC). The ROC plots the performance of a binary classifier system as a discrimination

rule is varied[50]. The AUC is a commonly used point-measure of the accuracy of the test, where 1 is the maximum possible value.

6.3.1 Experiment 1

Using the full set of features, the classifier yielded an AUC of 0.752 ± 0.055 (AUC$\pm$ SE), and the ROC curves are shown in both a continuous and stepwise manner in Figure 6.3.1. This result shows that SLIC generated supervoxels can be used to derive features for the successful classification of breast lesions. While the performance of this system is not as high as some reported in literature [6, 19], it is using only supervoxel-defined features and in this way the result is very promising.



(a) Empirical (raw) ROC curve.

(b) Smoothed ROC curve.

Figure 6.3.1: ROC curves for classification experiment.
b) shows the smoothed version of a) for the classification experiment when the full feature set was used. The AUC $= 0.752 \pm 0.055$.

The importance scores for each of the features used in the classifier are shown in Figure 6.3.2. The majority of features show an importance score greater than the R1 random feature, which is taken to be the baseline. No DW-MRI features rank below the baseline, while several DCE-MRI features do.

The feature scores after subtracting the baseline importance of the R1 random features are shown in Figure 6.3.3.

The ranking of importance scores shows that features derived from the DW-MRI modality appear to be of more importance to the classifier than those from the DCE-MRI modality. The ten most important scores for the classifier are listed in Table 6.1. The top ten importance scores are dominated by DW-MRI based features, and the large cluster size represents five of the top ten. The most important DCE-MRI feature is ranked at #21, and is feature $D32$ - the Agliozzo Model parameter A, for medium sized clusters.

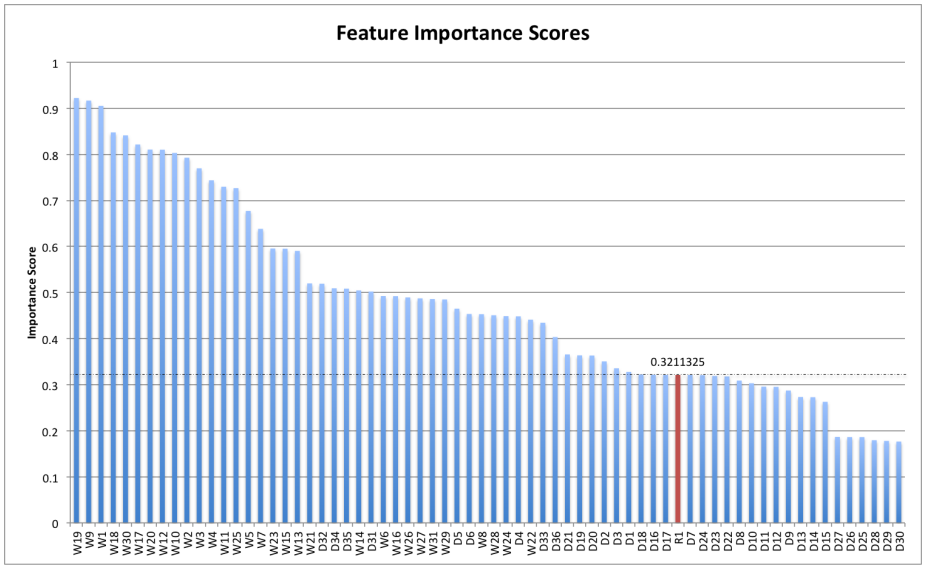


Figure 6.3.2: Importance Scores Ranked.
Features are ranked in descending order, and the randomly generated feature, R1 is marked in red. Features that have a greater importance score than R1 are deemed useful features.

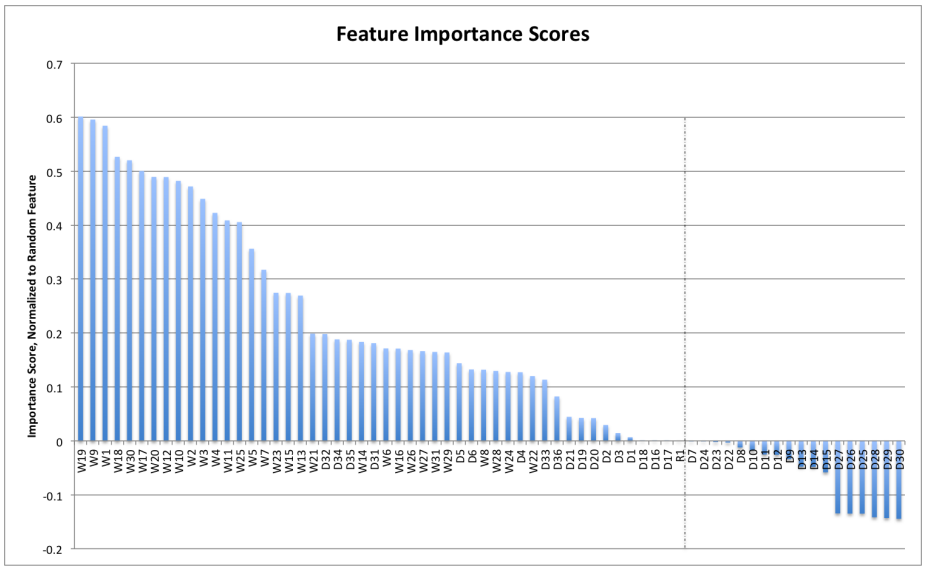


Figure 6.3.3: Importance scores with respect to R1 random feature.
In this plot, the vertical axis are the scores after subtracting the score of the random feature R1.

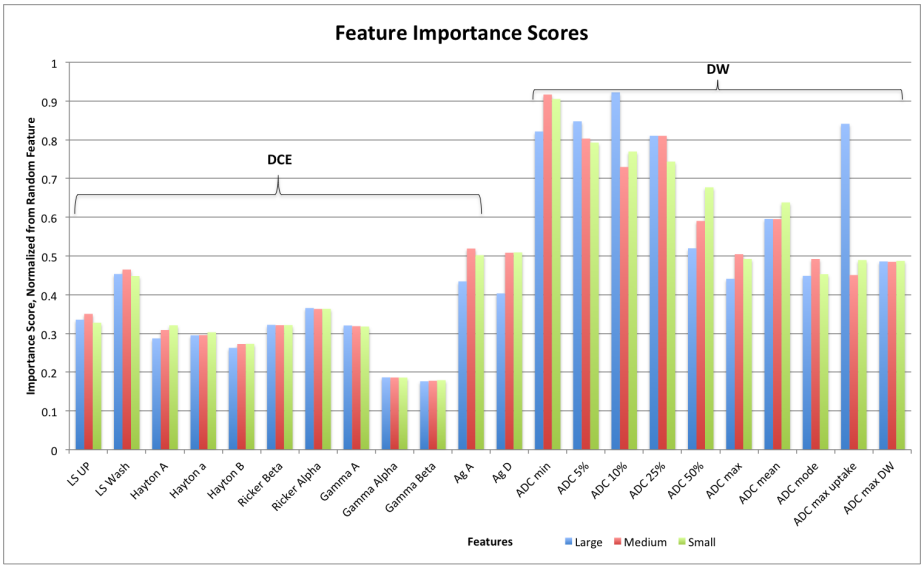


Figure 6.3.4: Importance Scores grouped by feature and size

Table 6.1: Top 10 Importance Scores

Rank	Feature
1	10th percentile ADC in large clusters ($W19$)
2	Minimum ADC in medium clusters ($W9$)
3	Minimum ADC in small clusters ($W1$)
4	5th percentile ADC in large clusters ($W18$)
5	ADC in cluster of max contrast uptake, large clusters ($W30$)
6	Minimum ADC in large clusters ($W17$)
7	25th percentile ADC in large clusters ($W20$)
8	25th percentile ADC in medium clusters ($W12$)
9	5th percentile ADC in medium clusters ($W10$)
10	5th percentile ADC in small clusters ($W2$)

In Figure 6.3.6 the importance scores are grouped by features and size. Grouping the importance scores by size however does not show any clear trend however, as only rarely is there a significant difference between sizes. The feature where the size seems to have played the largest role is the ADC in the cluster of maximum contrast uptake. In this feature, the large cluster model produced drastically higher importance scores.

6.3.2 Experiment 2

Running the classification experiment while varying the *mtry* parameter produced stable results (which concurs with the available literature). These results are shown in Table 6.2. This result shows that RF is a classification method where little tuning is required to attain the acceptable performance, as the variation of the main tuning parameter *mtry* shows little effect.

Table 6.2: Classification Performance vs *mtry* values

mtry	Mean AUC
2	0.7411
3	0.7518
4	0.7443
5	0.7438
10	0.7375
20	0.7339

Table 6.3: Classification performance vs *ntree* values

ntree	Mean AUC
500	0.7508
2000	0.7518
3000	0.7448
4000	0.7446
10000	0.7588

6.3.3 Experiment 3

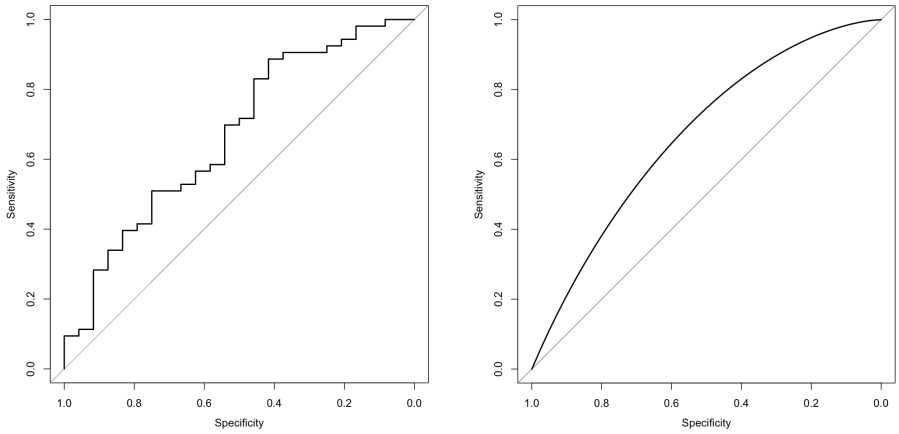
Similarly, as expected the classification results are stable with respect to the number of trees generated, *ntree*. The mean AUC with respect to number of trees is shown in Table 6.3. This result again demonstrates that little tuning is required for the RF classification algorithm to attain acceptable performance, as the performance was stable across the range of *ntree*.

6.3.4 Experiment 4

Running the classification experiment with only the DCE-MRI derived feature set produces poorer results, as expected. This trial yielded an AUC of 0.670 ± 0.063. The ROC curves are shown in Figure 6.3.5.

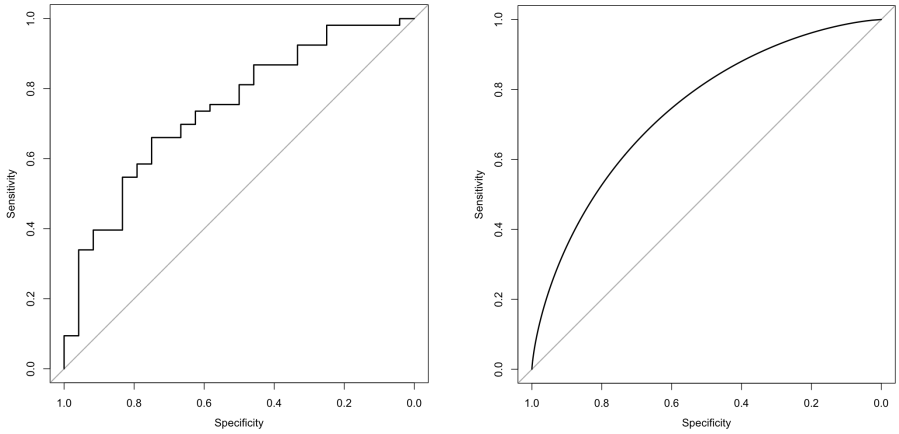
6.3.5 Experiment 5

Similarly running the classification with only the DW-MRI derived feature set produces poorer results than the full list of features, but only marginally. This trial yielded an AUC of 0.743 ± 0.056. The ROC curves are shown in 6.3.6. Comparing this result to that of [19] confirms that the cluster-defined features are indeed the most valuable for use in a classification experiment while using the DW-MRI modality. This fits the hypothesis that cluster-based features will have lower noise as they are less impacted by noisy voxels or errors in registration.



(a) Empirical (b) Continuous

Figure 6.3.5: ROC curves for classification using only DCE-MRI features. Using only the DCE-MRI feature set, the AUC $= 0.670 \pm 0.063$.



(a) Empirical (b) Continuous

Figure 6.3.6: ROC curves for classification using only DW-MRI features. Using only the DW-MRI feature set, the AUC $= 0.743 \pm 0.056$.

Chapter 7

Summary and Conclusion

This chapter reviews the work presented in this thesis, summarizes contributions and findings, details limitations in the work, and proposes future avenues for research.

7.1 Thesis Summary

Chapter 1 This chapter provided context to the project's work, and described the current limitations and avenues of research with respect to computer aided diagnosis (CAD) systems for breast cancer. Specifically, it noted that CAD systems rely on hierarchical segmentations (from large scale breast-air boundary to inter-lesion clusters) and methods for clustering voxels that share similar characteristics. The chapter also describes the major aim of the thesis:

> "to explore the use of the SLIC supervoxel algorithm within the context of an automatic diagnosis system for breast cancer".

and objectives:

1. To implement SLIC for the generation of supervoxels in breast MRI
2. To develop and evaluate a supervoxel-based method to segment the breast-air boundary.
3. To develop and evaluate supervoxel-based features for classifying breast lesions as benign or malignant.

Chapter 2 This chapter presented Simple Linear Iterative Clustering (SLIC). First, an overview of the uses of superpixels and supervoxels is provided, and the reasons why SLIC was chosen for this project are noted. Specifically, SLIC has the ability to be extended into 3D, can be tailored quickly for use in MRI, is tuneable with only a single parameter m, provides control over the number of supervoxels created, and is computationally efficient. Next, the algorithm is described step-by-step. Finally, sample segmentations of full volumes are presented with ranging compactness levels and levels of complexity reduction.

Chapter 3 This chapter presented a well-known method used to segment the breast-air boundary in MRI, outlining its method of action, as well as noting situations in which the algorithm fails. Next, a supervoxel based variation for segmenting the breast-air boundary is proposed. This algorithm utilizes the SLIC algorithm to create supervoxels, from which a region adjacency graph is defined, and finally a graph-cuts approach is used to segment the volume. Several sample segmentations were shown, and the strengths and weaknesses of the proposed algorithm were discussed.

Chapter 4 This chapter presented an adaptation of the SLIC algorithm for use in a multi-modal segmentation procedure inside of radiologist defined volumes of interest. First, the changes to the algorithm are discussed and the input volumes are detailed. Specifically, the algorithm is adapted to use four dynamically contrast-enhanced (DCE-MRI) volumes, where the first volume is acquired pre-contrast injection and the subsequent three are acquired after contrast is administered. Sample segmentations of radiologist defined lesions are presented.

Chapter 5 This chapter presented several quantitative features for characterising breast lesions that are derived from the SLIC supervoxel algorithm presented in Chapter 4. These features are derived from two MR modalities: diffusion weighted (DW-MRI) and DCE-MRI.

Chapter 6 This chapter presented the classification experiments that were used to evaluate the efficacy of the features presented in Chapter 5 classifying lesions as benign or malignant. First, the chapter begins with a review of the classification method used: Random Forests (RF). A description of the algorithm is provided, as well as the reasoning behind it's use. The results show that using the combination of both DW- and DCE-MRI the classifier yields the best classification performance of 0.752 ± 0.055 (AUC$\pm$SE).

7.2 Key Contributions and Findings

- An automated method to segment MRI volumes into spatially coherent supervoxels based on the SLIC algorithm. The method makes use of both the spatial coordinates of the voxel and it's signal intensity vector during the clustering process. The method takes only one tuning parameter aside from the desired number of supervoxels as input. This tuning parameter can be used to control the compactness of the resulting supervoxels. The empirical results show that the algorithm creates supervoxels that adhere well to the image boundaries.
- An automated method to segment breast MRI volumes to create a mask that can be used to exclude the background of air in the volume. This is an important preprocessing in a CAD system used to ensure analysis is performed solely on breast tissue. The proposed method is based on the graph-cuts algorithm, where the graph is created on the basis of supervoxels. This method, because it uses the supervoxels as a way of reducing

complexity in the volume, will likely become more valuable if the size of the volumes increases.

- An automated method to create supervoxels inside arbitrary volumes of interest, created on the basis of the DCE volume time series. Clustering the voxels based on their enhancement pattern leads to supervoxels that contain similar tissue. This SLIC-based method can be tuned by altering the size of the supervoxels and their compactness.
- A set of quantitative features, derived from the partitioning of a suspicious lesion into SLIC supervoxels.
- Evidence that the proposed supervoxel-based features have efficacy for classifying suspicious lesions as either benign or malignant using a Random Forest classifier. The experimental results demonstrate an AUC of 0.752 ± 0.055.
- Classification results using only DW-MRI concur with the findings of [19],wherein a similar performance was attained using only the cluster-based descriptors. In [19], of the top 10 most important features for this scenario, four were cluster based. Given that DW-MRI volumes can be particularly noisy, and are often affected by registration errors cluster-based features are useful because they are less affected by these factors.

7.3 Conclusion

This thesis demonstrates that SLIC Supervoxels can be applied in a breast MRI CAD system for both automated segmentation tasks and feature extraction. This thesis shows that SLIC can be used to generate segmentations across a range of purposes, from large-scale segmentation tasks such as the breast-air boundary, to small-scale tasks such as creating clusters of similar tissue inside a VOI. SLIC can be used in a single-modality situation, or in a multimodal situation. Useful information can also be extracted on the basis of supervoxels for use in a classification system.

SLIC can be used as a pre-processing stage in order to reduce the level of complexity, or memory requirements for other processes. In this work, this was demonstrated through the segmentation of the breast-air boundary where it reduced the memory requirements for a popular graph-cuts segmentation algorithm.

Lesions can be classified as either benign or malignant on the basis of features defined on the supervoxel level, which opens up the possibility for further research into supervoxel-based descriptors of lesions.

With the increasing number of MRI screenings for breast cancer, as advocated for by the American Cancer Society, the need for fast and accurate CAD systems to aid the radiologists in this increasingly complex and time-consuming task is large. This work has contributed several proof-of-concept CAD algorithms using an efficient, and flexible algorithm based on SLIC supervoxels.

7.4 Limitations and Future Work

One of the largest limitations of this work is that it did not consider the task of automatically locating and delineating suspicious lesions in the breast volume. This task is central to an effective and useful CAD system. This task has been addressed in the literature using mean-shift [15, 19] and it may be possible to implement SLIC in a similar manner. This could be an interesting avenue of future work, because SLIC is known to be computationally efficient and this is an important aspect for any future commercial product. This would require further research into methods that could be used to join supervoxels. It is likely that the same large-scale segmentation used to segment the breast-air boundary could be used in this process.

Another limitation of the current work is that it has only been tested on a controlled set of MRI volumes. The data that was used was acquired using the same acquisition protocol, and does not suffer from major artefacts, registration issues, or other image quality problems. While the SLIC implementation here was created in a flexible manner, compatibility or implementation issues could arise. For example, further preprocessing scripts ensuring that no negative intensity values exist in the volume may be necessary.

Aside from segmentation of the breast-air boundary, another large-scale segmentation task is also frequently performed in a breast MRI CAD system: removing the chest cavity. This task was not addressed in this work. The same large-scale segmentation that was used to segment the breast-air boundary could be used in this task as well, although the proposed method may need tuning. For example, the chest wall-breast tissue boundary is not marked by as large a gradient as the breast-air boundary, and this can cause difficulties. A graph-cuts approach also has the potential to succeed in this scenario, although more specific object and background definitions would likely be required. This task could also be approached using prior knowledge of the shape of the chest cavity and its location in the axially-acquired volumes as is done in [35]. Using this information, specific seed locations could be created. Thresholding of supervoxel centre intensities using Otsu's method may even provide useful results. Alternatively, multi-class graph-cuts could be investigated or a sequential graph-cuts approach where another cut is formed on the volume after the air is masked.

An interesting area for future development could be in the creation of features derived from the shape of supervoxels, or in measures that use not only one supervoxel to derive information but use groupings of supervoxels. An example of this could be useful in small-scale segmentations inside VOIs, where signs of malignancy such as centripetal or centrifugal enhancement[18] could be described on a supervoxel basis. This could be achieved by creating a distance-from-VOI-centre measurement, and comparing each supervoxel's mean intensity and distance to centre. Using supervoxels aggregates voxel data and reduces variation in the resulting features could lead to increased classifier performance.

Bibliography

[1] N. C. Institute, "NCI SEER Stat Fact Sheets : Breast Cancer," United States National Cancer Institute, Bethesda,MD,USA, Tech. Rep., 2013. [Online]. Available: http://seer.cancer.gov/statfacts/html/breast.html

[2] American Cancer Society, "Types of breast cancers," pp. 1–3, 2013. [Online]. Available: http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-breast-cancer-types

[3] M. D. Dorrius, M. C. Jansen-van der Weide, P. M. a. van Ooijen, R. M. Pijnappel, M. Oudkerk, M. C. J.-v. D. Weide, and P. M. A. V. Ooijen, "Computer-aided detection in breast MRI: a systematic review and meta-analysis." *European radiology*, vol. 21, no. 8, pp. 1600–8, Aug. 2011. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3128262&tool=pmcentrez&rendertype=abstract

[4] R. Achanta, A. Shaji, K. Smith, and A. Lucchi, "SLIC superpixels," EPFL, Laussanne, Tech. Rep. June, 2010. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.165.8269&rep=rep1&type=pdf

[5] ——, "SLIC Superpixels Compared to State-of-the-Art Superpixel Methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2281, 2012.

[6] S. Agliozzo, M. D. Luca, C. Bracco, M. De Luca, a. Vignati, V. Giannini, L. Martincich, L. a. Carbonaro, a. Bert, F. Sardanelli, and D. Regge, "Computer-aided diagnosis for dynamic contrast-enhanced breast MRI of mass-like lesions using a multiparametric model combining a selection of morphological, kinetic, and spatiotemporal features." *Medical physics*, vol. 39, no. 4, pp. 1704–15, Apr. 2012. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22482596http://scitation.aip.org/content/aapm/journal/medphys/39/4/10.1118/1.3691178

[7] C. Dromain, B. Boyer, R. Ferré, S. Canale, S. Delaloge, and C. Balleyguier, "Computed-aided diagnosis (CAD) in the detection of breast cancer." *European journal of radiology*, vol. 82, no. 3, pp. 417–23, Mar. 2013. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22939365

[8] D. Saslow, C. Boetes, W. Burke, S. Harms, M. O. Leach, C. D. Lehman, E. Morris, E. Pisano, M. Schnall, S. Sener, R. a. Smith, E. Warner, M. Yaffe, K. S. Andrews, and C. a. Russell, "American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography."

CA: a cancer journal for clinicians, vol. 57, no. 2, pp. 75–89, 2007. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/17392385

[9] K. Mcpherson, C. M. Steel, and J. M. Dixon, "Breast cancer - epidemiology, risk factors and genetics," *British Medical Journal*, vol. 321, no. September, 2000.

[10] "World Cancer Report 2008," World Health Organization - International Agency for Research on Cancer, Lyon, Tech. Rep., 2008.

[11] D. R. Youlden, S. M. Cramb, N. a. M. Dunn, J. M. Muller, C. M. Pyke, and P. D. Baade, "The descriptive epidemiology of female breast cancer: an international comparison of screening, incidence, survival and mortality." *Cancer epidemiology*, vol. 36, no. 3, pp. 237–48, Jun. 2012. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22459198

[12] S. Sinha and U. Sinha, "Recent advances in breast MRI and MRS." *NMR in biomedicine*, vol. 22, no. 1, pp. 3–16, Jan. 2009. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18654998

[13] U. o. P. M. Center, "Sample Mammogram," 2014. [Online]. Available: http://www.upmc.com/services/imaging/services/women/services/breast/pages/mammogram.aspx

[14] A. C. of Radiologists, "ACR - Bi-rads," Tech. Rep., 2003.

[15] D. McClymont, A. Mehnert, A. Trakic, D. Kennedy, and S. Crozier, "Fully automatic lesion segmentation in breast MRI using mean-shift and graph-cuts on a region adjacency graph," *Journal of Magnetic Resonance Imaging*, Oct. 2013. [Online]. Available: http://doi.wiley.com/10.1002/jmri.24229

[16] M. J. Stoutjesdijk, J. Veltman, H. Huisman, N. Karssemeijer, J. O. Barentsz, J. G. Blickman, and C. Boetes, "Automated analysis of contrast enhancement in breast MRI lesions using mean shift clustering for ROI selection." *Journal of magnetic resonance imaging : JMRI*, vol. 26, no. 3, pp. 606–14, Sep. 2007. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/17729367

[17] D. Newell, K. Nie, J.-H. Chen, C.-C. Hsu, H. J. Yu, O. Nalcioglu, and M.-Y. Su, "Selection of diagnostic features on breast MRI to differentiate between malignant and benign lesions using computer-aided diagnosis: differences in lesions presenting as mass and non-mass-like enhancement." *European radiology*, vol. 20, no. 4, pp. 771–81, Apr. 2010. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2835636&tool=pmcentrez&rendertype=abstract

[18] W. a. Kaiser, *Signs in Breast MR*. Springer Science & Business Medical, 2009.

[19] D. McClymont, "Computer assisted detection and characterization of breast cancer in MRI," Ph.D. dissertation, University of Queensland, Australia, 2014.

[20] A. Chan, A. A, and S. S. J. S. Nelson, "Simplified gamma-variate fitting of perfusion curves," *2004 2nd IEEE International Symposium on Biomedical Imaging: Macro to Nano (IEEE Cat No. 04EX821)*, vol. 2, no. 2, pp. 1067–1070, 2004. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1398726

[21] P. Hayton, M. Brady, L. Tarassenko, N. Moore, M. R. I. Unit, J. R. Hospital, H. Way, and O. Ox, "Analysis of dynamic MR breast images using a model of contrast enhancement." *Medical image analysis*, vol. 1, no. 3, pp. 207–24, Apr. 1997. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/9873907

[22] A. Mehnert, M. Wildermoth, S. Crozier, D. Kennedy, and E. Bengtsson, "Two Non-linear Parametric Models of Contrast Enhancement for DCE-MRI of the Breast Amenable to Fitting Using Linear Least Squares," *2010 International Conference on Digital Image Computing: Techniques and Applications*, pp. 611–616, Dec. 2010. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5692629

[23] W. Ricker, "Handbook of computations for biological statistics of fish populations," *Bull. Fish. Res. Bd. Can*, vol. 119, pp. 1–300, 1958.

[24] O. Schabenberger, *Contemporary Statistical models for the plant and soil sciences*. CRC Press, 2002.

[25] X. Ren and J. Malik, "Learning a classification model for segmentation," *Proceedings Ninth IEEE International Conference on Computer Vision*, vol. 1, no. c, pp. 10–17 vol.1, 2003. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1238308

[26] The MathWorks inc, "MATLAB 2013b," 2013.

[27] The R Project, "The R project for Statistical Computing," 2014. [Online]. Available: www.r-project.org

[28] J. Cheng, J. Liu, Y. Xu, F. Yin, D. W. K. Wong, B.-H. Lee, C. Cheung, T. Aung, and T. Y. Wong, "Superpixel classification for initialization in model based optic disc segmentation." *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, vol. 2012, pp. 1450–3, Jan. 2012. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/23366174

[29] K.-S. Kim, D. Zhang, M.-C. Kang, and S.-J. Ko, "Improved Simple Linear Iterative Clustering Superpixels," in *2013 IEEE 17th International Symposium on Consumer Electronics(ISCE)*. IEEE, 2013, pp. 259–260.

[30] A. Vedaldi and S. Soatto, "Quick Shift and Kernel Methods for Mode Seeking," *Computer Vision ECCV 2008*, 2008. [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-540-88693-8_52

[31] A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi, "TurboPixels: fast superpixels using geometric flows." *IEEE transactions on pattern analysis and machine*

intelligence, vol. 31, no. 12, pp. 2290–7, Dec. 2009. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19834148

[32] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient Graph-Based Image Segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, Sep. 2004. [Online]. Available: http://link.springer.com/10.1023/B:VISI.0000022288.19776.77

[33] N. Verma, M. C. Cowperthwaite, and M. K. Markey, "Superpixels in brain MR image analysis." *Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, vol. 2013, pp. 1077–80, Jan. 2013. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/24109878

[34] A. Lucchi, K. Smith, R. Achanta, G. Knott, and P. Fua, "Supervoxel-based segmentation of mitochondria in em image stacks with learned shape features." *IEEE transactions on medical imaging*, vol. 31, no. 2, pp. 474–86, Feb. 2012. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/21997252

[35] P. Hayton, "Analysis of contrast-enhanced breast MRI," Ph.D. dissertation, Oxford University, Oxford, UK, 1998.

[36] J. Nagi, S. A. Kareem, F. Nagi, and S. K. Ahmed, "Detection Using Digital Mammograms," *IEEE EMBS Conference on Biomedical Engineering & Sciences*, no. December, pp. 87–92, 2010.

[37] M. a. Wirth and A. Stapinski, "Segmentation of the breast region in mammograms using active contours," in *Visual Communications and Image Processing 2003*, T. Ebrahimi and T. Sikora, Eds., Jun. 2003, pp. 1995–2006. [Online]. Available: http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=768088

[38] Y. Y. Boykov, "Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images," *Proceedings of the International Conference on Computer Vision*, no. July, pp. 105–112, 2001.

[39] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, third edit ed. CENGAGE Learning, 2008.

[40] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision." *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 9, pp. 1124–37, Sep. 2004. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15742889

[41] Y. Boykov and G. Funka-Lea, "Graph Cuts and Efficient N-D Image Segmentation," *International Journal of Computer Vision*, vol. 70, no. 2, pp. 109–131, Nov. 2006. [Online]. Available: http://link.springer.com/10.1007/s11263-006-7934-5

[42] O. Veksler, Y. Boykov, and P. Mehrani, "Superpixels and Supervoxels in an Energy Optimization Framework," *Computer Vision ECCV 2010,*, pp. 211–224, 2010.

[43] J. Stawiaski, E. Decenciére, J. S. And, and E. Decenciere, "Region Merging via Graph-Cuts," *Image Anal stereol*, vol. 27, pp. 39–45, 2008. [Online]. Available: http://hal.archives-ouvertes.fr/hal-00830726/

[44] E. Morris and L. Liberman, *Breast MRI: Diagnosis and Intervention*. Springer, 2005.

[45] M. Hatakenaka, H. Soeda, H. Yabuuchi, Y. Matsuo, T. Kamitani, Y. Oda, M. Tsuneyoshi, and H. Honda, "Apparent diffusion coefficients of breast tumors: clinical application." *Magnetic resonance in medical sciences : MRMS : an official journal of Japan Society of Magnetic Resonance in Medicine*, vol. 7, no. 1, pp. 23–9, Jan. 2008. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18460845

[46] H. Yabuuchi, Y. Matsuo, T. Okafuji, T. Kamitani, H. Soeda, T. Setoguchi, S. Sakai, M. Hatakenaka, M. Kubo, N. Sadanaga, H. Yamamoto, and H. Honda, "Enhanced mass on contrast-enhanced breast MR imaging: Lesion characterization using combination of dynamic contrast-enhanced and diffusion-weighted MR images." *Journal of magnetic resonance imaging : JMRI*, vol. 28, no. 5, pp. 1157–65, Nov. 2008. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18972357

[47] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001. [Online]. Available: http://link.springer.com/article/10.1023/A:1010933404324

[48] R. Díaz-Uriarte and S. Alvarez de Andrés, "Gene selection and classification of microarray data using random forest." *BMC bioinformatics*, vol. 7, p. 3, Jan. 2006. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1363357&tool=pmcentrez&rendertype=abstract

[49] W. G. Touw, J. R. Bayjanov, L. Overmars, L. Backus, J. Boekhorst, M. Wels, and S. a. F. T. van Hijum, "Data mining in the Life Sciences with Random Forest: a walk in the park or lost in the jungle?" *Briefings in bioinformatics*, vol. 14, no. 3, pp. 315–26, May 2013. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3659301&tool=pmcentrez&rendertype=abstract

[50] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. Springer, 2005. [Online]. Available: http://link.springer.com/content/pdf/10.1007/978-0-387-84858-7.pdf

[51] A. Bureau, J. Dupuis, K. Falls, K. L. Lunetta, B. Hayward, T. P. Keith, and P. Van Eerdewegh, "Identifying SNPs predictive of phenotype using random forests." *Genetic epidemiology*, vol. 28, no. 2, pp. 171–82, Feb. 2005. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15593090