UC Irvine

UC Irvine Electronic Theses and Dissertations

Title

Development of Segmentation Algorithms and Machine Learning Classification Methods for Characterization of Breast and Spine Tumors on MRI

Permalink

https://escholarship.org/uc/item/5r1552k7

Author

Wang, Xinxin

Publication Date

2017

Supplemental Material

https://escholarship.org/uc/item/5r1552k7#supplemental

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA,
IRVINE

Development of Segmentation Algorithms and Machine Learning Classification Methods for Characterization of Breast and Spine Tumors on MRI

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in Biomedical Engineering

by

Xinxin Wang

Thesis Committee:
Associate Professor Gultekin Gulsen, Chair
Professor Min-Ying Lydia Su
Professor Frithjof Kruggel

2017

© 2017 Xinxin Wang

Dedication

To

My parents and my friends

for their endless love and unconditional support

Table of Contents

LIST OF FIGURES v
LIST OF TABLES vi
ACKNOWLEGMENTS vii
ABSTRACT OF THE THESIS viii
Chapter 1 Introduction ..... - 1 -
1.1 Magnetic Resonance Imaging ..... - 1 -
1.2 Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) ..... - 2 -
Chapter 2 Mass and non-mass tumor segmentation in breast cancer ..... - 4 -
2.1 Introduction ..... - 4 -
2.2 Subjects and MRI Protocols ..... - 5 -
2.3 Breast quadrant segmentation ..... - 6 -
2.4 Methodology of breast tumor segmentation ..... - 9 -
2.4.1 Mass tumor segmentation ..... - 10 -
2.4.2 Non-mass tumor segmentation ..... - 11 -
2.5 Results ..... - 14 -
2.6 Summary and Discussion ..... - 20 -
Chapter 3 Spinal cord cancer segmentation ..... - 21 -
3.1 Introduction ..... - 21 -
3.2 Subjects and MRI Protocols ..... - 22 -
3.3 Methodology of spinal lesion segmentation ..... - 23 -
3.3.1 Registration from sagittal view to DCE axial sequence ..... - 23 -
3.3.2 Normalized Cut ..... - 25 -
3.3.3 Region growing ..... - 28 -
3.4 Results ..... - 29 -
3.5 Summary and Discussion ..... - 32 -
Chapter 4 Machine Learning ..... - 33 -
4.1 Introduction ..... - 33 -
4.2 Breast cancer machine learning ..... - 35 -
4.3 Spinal cancer machine learning ..... - 40 -
4.4 Summary and Discussion ..... - 45 -
Chapter 5 Summary ..... - 47 -

Reference ..... - 49 -
Appendix ..... - 51 -

LIST OF FIGURES

Figure 2.1 Comparison of mass tumor and non-mass-like tumor ..... - 4 -
Figure 2.3.1 Three views in 3D model of a breast and its fibroglandular tissue ..... - 7 -
Figure 2.3.2 Quadrant segmentation result of the case in Figure 2.3.1 ..... - 7 -
Figure 2.3.3 Three views in 3D model of a breast and the tumor ..... -8 -
Figure 2.3.4 Quadrant segmentation result of the case in Figure 2.3.3 ..... - 8 -
Figure 2.4.1 Flow chart of segmentation procedures for mass and non-mass tumors ..... - 9 -
Figure 2.4.2 Procedures for non-mass tumor segmentation ..... - 12 -
Figure 2.4.3 Main interface of the segmentation program ..... - 13 -
Figure 2.5.1 Example of four mass tumor segmentation results ..... - 14 -
Figure 2.5.2 Example of two non-mass tumor segmentation results ..... - 15 -
Figure 2.5.3 3D model of segmented non-mass tumor ..... - 16 -
Figure 3.3.1 Basic steps showing registration from sagittal view to DCE axial sequence - 25 -
Figure 3.3.2 An example picture using normalized cut algorithm ..... - 26 -
Figure 3.3.3 Lymphoma patient case using normalized cut algorithm ..... - 27 -
Figure 3.3.4 Lymphoma patient case using region growing algorithm ..... - 29 -
Figure 3.4.1 Chordoma Case ..... - 29 -
Figure 3.4.2 Giant Cell Tumor (GCT) Case ..... - 30 -
Figure 3.4.3 Granuloma Case ..... - 30 -
Figure 3.4.4 Lymphoma Case ..... - 30 -
Figure 3.4.5 Metastatic cancer Case ..... - 31 -
Figure 3.4.6 Myeloma Case ..... - 31 -
Figure 3.4.7 TB Case ..... - 31 -
Figure 4.2.1 ROC curve and AUC among three tumor subtypes ..... - 40 -
Figure 4.3.1 Examples of three maps corresponding to three definitions ..... - 41 -
Figure 4.3.2 Brief introduction of 50 texture features ..... - 42 -
Figure 4.3.3 Classification Error and predictors' importance estimates between GCT and Granuloma ..... - 43 -
Figure 4.3.4 Classification Error and predictors' importance estimates between GCT and TB ..... - 43 -
Figure 4.3.5 Classification Error and predictors' importance estimates between TB and Granuloma ..... - 44 -
Figure 4.3.6 ROC curve of classification result among three types ..... - 45 -

LIST OF TABLES

Table 2.5.1 Examples of comparison between pre and post edited lesion areas ..... - 17 -

Table 4.2.1 ANOVA of 11 morphology features..... - 37 -

Table 4.3.1 AUC result of 10 different combination among three types ..... - 45 -

ACKNOWLEGMENTS

Firstly, I would like to express my sincere gratitude to my advisor, Prof. Lydia Su, for her continuous guidance and support during my research at the Center for Functional Onco-Imaging. Her patience, encouragement and inspiration always helped me to tackle with challenges during research and guided me to find out the best solution.

Besides my advisor, I want to express my appreciation to my clinical supervisor, Dr. Jeon-Hor Chen. He has been providing me with continuous help in medical-related problems and giving me insights from a doctor's prospective. His strictness in researches gave me precious lessons and invaluable experience.

I also want to thank the members of my thesis committee, Prof. Gultekin Gulsen and Dr. Frithjof Kruggel. Without their support and insightful comments to my thesis, I won't be able to complete my thesis from various perspectives. Also, what they taught me during lectures in previous quarters were priceless treasure which let me make progress and be more confident doing research in this field.

Last but not least, I would like to say thank you to my lab mates for giving me a hand when I was encountering difficulties in research. I would like to appreciate my friends for being with me in the past years and always make me feel being loved. And of course, I would like to thank my parents for their love and care, without which my Master's research and thesis wouldn't have been possible.

ABSTRACT OF THE THESIS

Application of MRI image processing techniques in breast tumor and spinal tumor segmentation

By

Xinxin Wang

Master of Science in Biomedical Engineering

University of California, Irvine, 2017

Associate Professor Gultekin Gulsen, Chair

As the second most leading cause of cancer death in women, breast cancer has attracted wide attention in MRI studies. However, the segmentation of non-mass lesions in breast cancer MRI images remains challenging due to their comparatively high variations in kinetic characteristics and typical morphological parameters. Meanwhile, the segmentation of spinal tumors in axial and sagittal views has been tough task because of the complexity to determine lesion area in one single view.

In this thesis, we proposed a method for automatic non-mass tumor segmentation, which was threshold oriented and based on region growing algorithm. It turned out that the segmentation results were satisfactory where only less than 5% pixels compared to the total number of segmented tumor pixels were corrected. We also introduced a method for spinal tumor segmentation to determine the location of lesions in spinal cord cancers. Normalized cut and region growing were applied in this study so to locate spinal tumors. Seven types of

spinal cancers were analyzed and the overall behavior of the method was good in most cases except for some cases due to influence of unclear boundaries of the lesion areas. Finally, we applied a machine learning based method to differentiate three subtypes of breast cancer tumors and three types of spinal cancers. ANOVA and random forest algorithm was utilized to select the most significant features and cross-validation-based logistic regression was then used in this step in order to build a classifier. The performance of the classifier can be improved if more cases are included and affecting factors like biomarkers in spinal cancers can be rejected from the classification process.

Chapter 1 Introduction

1.1 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is known as an imaging technique used to observe the anatomy of the human body. It is a safe and noninvasive way to create detailed pictures of human organ and tissue. For safety, it means that compared with many other imaging techniques, it uses radio wave instead of ionization radiation to process imaging and for noninvasive, it implies that using this technique doesn't have to have any wounds on or any instruments insert into the patient's body.

The principle of MRI is that radio frequency energy can be absorbed and emitted by certain atomic nuclei located in an external powerful magnetic field. The most commonly used atoms in modern hospitals are hydrogen atoms. They are used to generate a detectable radio-frequency signal received by scanners in close proximity to the anatomy being examined. Image contrast is provided by the varying relaxation times between different tissue types. [1.1]

MRI is a powerful medical diagnostic tool, which can be applied to the detailed imaging of a variety of soft tissues, in particular the brain, in a wide range of clinical applications:

Cardiology – Cardiac MRI can create both moving and still pictures of the patient's heart. When patient occurs with heart problems, for example pericarditis and valve disease, cardiac MRI is a common way for the clinical cardiac doctors to access and diagnosis different diseases and conditions.

Cancer – Lesion area shows bright feature under MRI scanning, comparing with the dark background of the other tissue, it is quite easier for the doctor to detect and aim at the target area.

Clinical neurology – In this area, the most common used technique is functional MRI (fMRI), which is not quantitative but results have proved reliable and reproducible and the study is also quick to perform to monitor the condition of patient. Clinical applications can be made in epilepsy surgery, the study of schizophrenia, etc. [1.2]

Soft tissue damage - MRI is sensitive to changes in cartilage and bone structure resulting from injury, disease, or aging. It can detect herniated discs, pinched nerves, spinal tumors, spinal cord compression, and fractures.

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

Dynamic contrast-enhanced MR imaging is a noninvasive technique measuring the kinetics of the distribution of contrast agents within tissues and acquiring serial MRI images before, during and after the distribution of contrast agent. Prevalently used contrast agent includes exogenous gadolinium (Gd)-based contrast agents, the rapid development of which in the 1980s opened a new era for imaging of tumors and vascular systems. Different from traditional MRI, DCE-MRI enables a comprehensive depiction of the wash-in and wash-out contrast kinetics within tumors through which the nature of the bulk tissue properties can be obtained, instead of a simple shot of enhancement at a time point. Using MR contrast agents of different molecular weights, features like the vascular permeability of tumor models and extravascular-extracellular space can be assessed. Using multi-temporal scan and producing continuous dynamic images, DCE-MRI is able to more objectively reflect the

enhanced features of lesions after post-processing, in which sense providing pathophysiological characteristics outside the anatomical structure of the lesion like other functional imaging techniques.

A relatively high temporal resolution was usually used in early DCE-MRI studies, while different post-contrast frames obtained at different times can be used to find the optimal frame showing the best signal contrast between normal tissues and lesions. [1.3] Currently, DCE-MRI is used to characterize masses, stage tumors, and noninvasively monitor therapy. There have been a number of limitations of DCE-MRI despite its prevalent application in clinical situations, which include the overlap between benign and malignant tissues, difficulty in successfully treating microscopic diseases, and the instability in values of enhancement pattern with predictive ability concerning clinical outcomes. Future studies aiming standardizing DCE-MRI acquisition, analysis, and reporting methods will make wider dissemination of this promising imaging technique possible. [1.4, 1.5]

In this thesis, both breast tumors segmentation in **Chapter 2** and spinal cancer image processing in **Chapter 3** were based on DCE-MRI images. All of the feature extraction, selection and machine learning algorithms in **Chapter 4** of this study were applied on contrast-enhanced images.

Chapter 2 Mass and non-mass tumor segmentation in breast cancer

2.1 Introduction

Breast cancer is the second most leading cause of cancer death in women only after lung cancer. Breast cancer remains a major public health problem. The incidence is rising in most countries and is projected to rise further over the next 20 years despite current efforts to prevent the disease. The increased incidence is not surprising since there has been, in most countries, an increase in numbers of women with major breast cancer risk factors, including lower age of menarche, late age of first pregnancy, fewer pregnancies, shorter or no periods of breastfeeding, and a later menopause. Other risk factors which add to the burden of breast cancer are the increase in obesity, alcohol consumption, inactivity, and hormone replacement therapy (HRT). [2.1]

With the improved technology in imaging, automatic and quantitative analysis of breast cancer may provide clinically important information for diagnosis and treatment planning. Breast tumor segmentation has been a mature technique for mass tumors that have clear boundaries and uniform shapes, but the segmentation for non-mass-like enhancement lesions remain both qualitatively as well as quantitatively difficult to analyze.

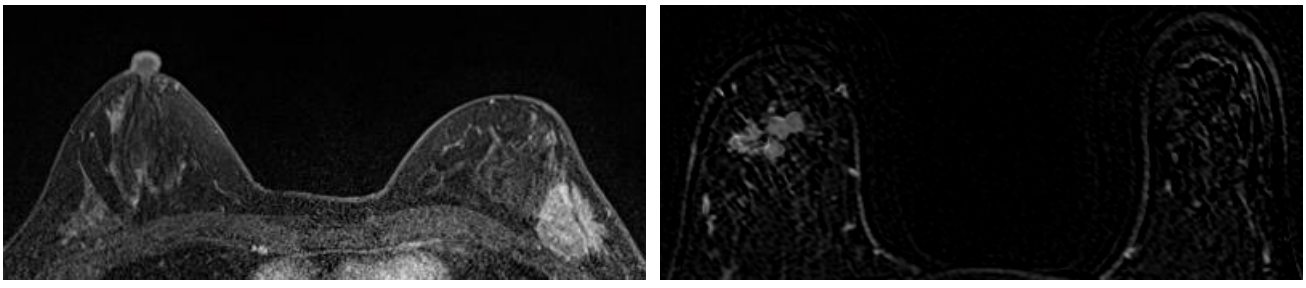


Figure 2.1 Comparison of mass tumor (left) and non-mass-like tumor (right).

While non-mass tumors have heterogeneous appearances in breast MRI, they also have comparatively high variations in kinetic characteristics and typical morphological parameters and lower reported specificity and sensitivity than mass-tumors. Compared to mass tumors, non-masses have no well-defined and blurred tumor borders and a kinetic behavior that is not easily generalizable and thus discriminative for malignant and benign non-masses. Non-mass tumor segmentations were either based on manually drawn ROI or semi-automatic segmentation in most of the previous study. [2.2 – 2.4]

The goal of this study is to develop a robust and reliable segmentation method for irregular non-mass tumors, which is based on threshold oriented region growing. Threshold in each case was determined by the distribution of the fibroglandular tissue and the fibroglandular tissue segmentation in this study was performed by using an automatic template-based method [2.5] After automatic segmentation, an experienced radiologist was asked to review all the non-mass-like cases and verify the quality of the segmentation.

2.2 Subjects and MRI Protocols

122 patients with breast cancer were analyzed in this retrospective study. The ages of the patients were from 22 to 75 and the mean age is 48.5 y/o. Among all 122 cases, 94 of them could be categorized by an experienced radiologist into three types: triple negative (14 cases range 37-64, mean age 50.3 y/o), HER2-positive (29 cases range 25-70, mean age 50.7 y/o), ER-positive, HER2-negative (51 cases range 28-69, mean age 47.9 y/o).

The MRI studies were performed using a Siemens 1.5T system. The imaging sequences included axial T1WI and axial T2WI were acquired using fast spin echo. Each of the sequence contains 96 images with 0.625* 0.625* 1.5 (mm3) voxel size.

The contrast agents, 0.1 [mmol/kg] Gd-DTPA, was injected after one set of pre-contrast images was acquired, by using an Ulrich power injector at a rate of 2 ml/s followed by 20 cc saline flush at the same rate. A total of 12 frames were acquired, so the total DCE-MRI acquisition time period ranged from 120 to 168 seconds. Depending on the abnormal area that needed to be covered and the choice of sequences to acquire different views, the total imaging time was approximately 25-30 minutes.

2.3 Breast quadrant segmentation

Quadrant breast segmentation is a method to separate a breast into four quadrants for quantitative measurements of the quadrant breast volume (BV) and density. The principle of this study is to find two perpendicular planes intersecting at the nipple or the nipple-centroid line as a standard to segment one breast into four relatively equal parts, inner lower, inner upper, outer lower and outer upper. After the BV, the fibroglandular tissue volume, and the percent density (PD) were calculated, the relationship between cancer occurrence rates and the quantitative breast density could be further investigated. [2.6]

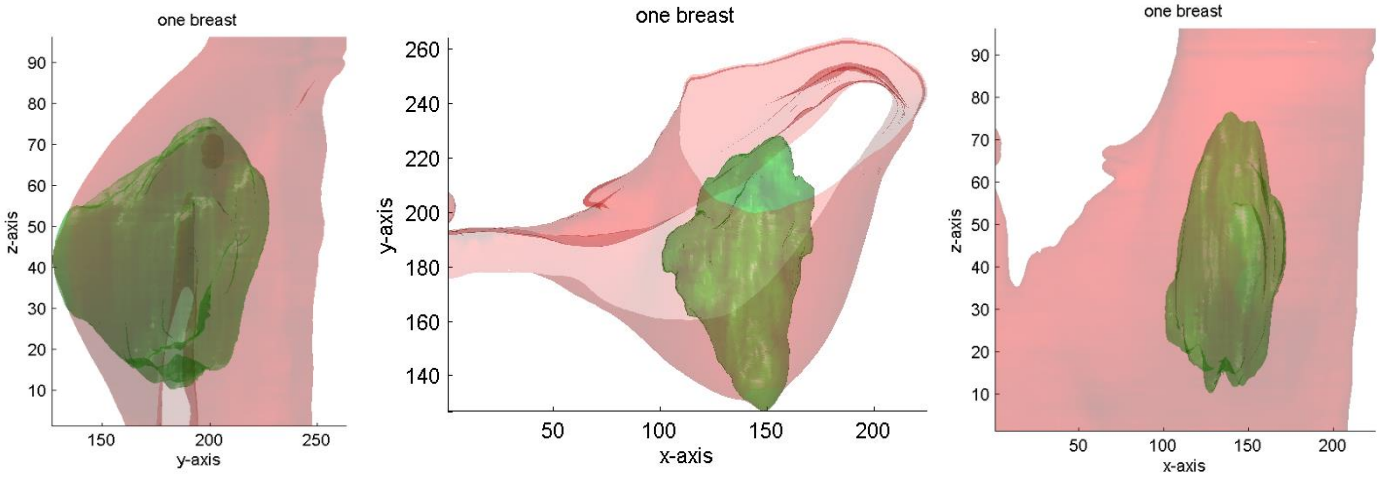


Figure 2.3.1 Three views in 3D model of a breast and its fibroglandular tissue, where pink area represents breast and green area represents fibroglandular tissues.

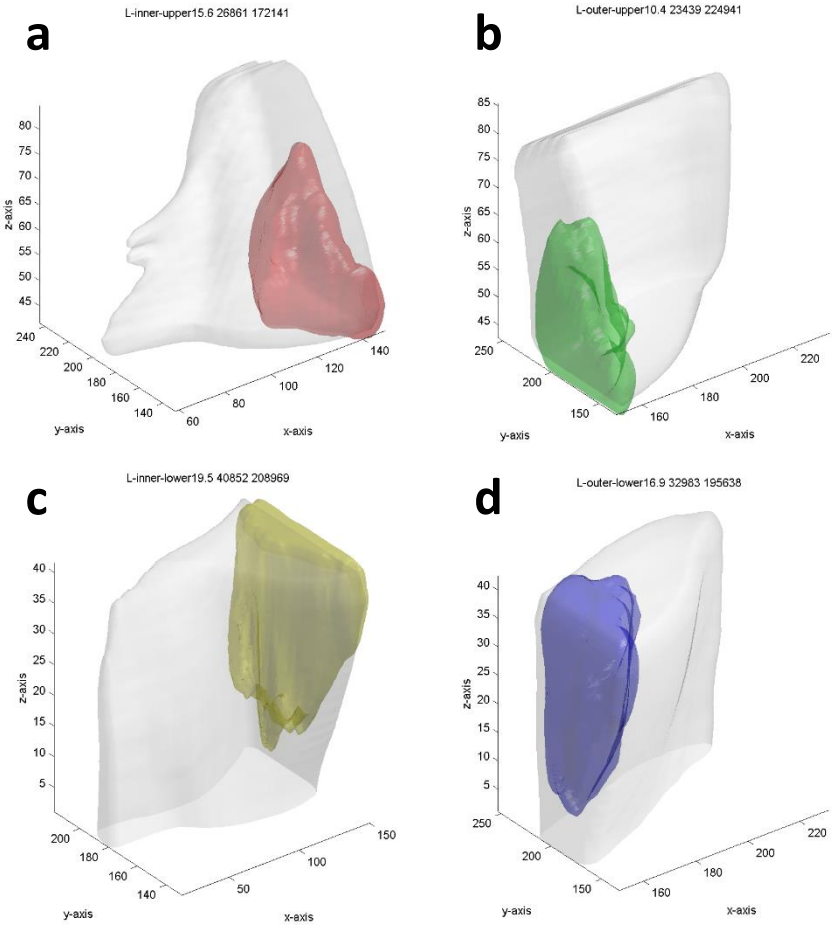


Figure 2.3.2 Quadrant segmentation result of the case in **Figure 2.3.1. (a), (b), (c), (d)** stands for the inner upper, outer upper, inner lower and outer lower quadrant of the breast, respectively.

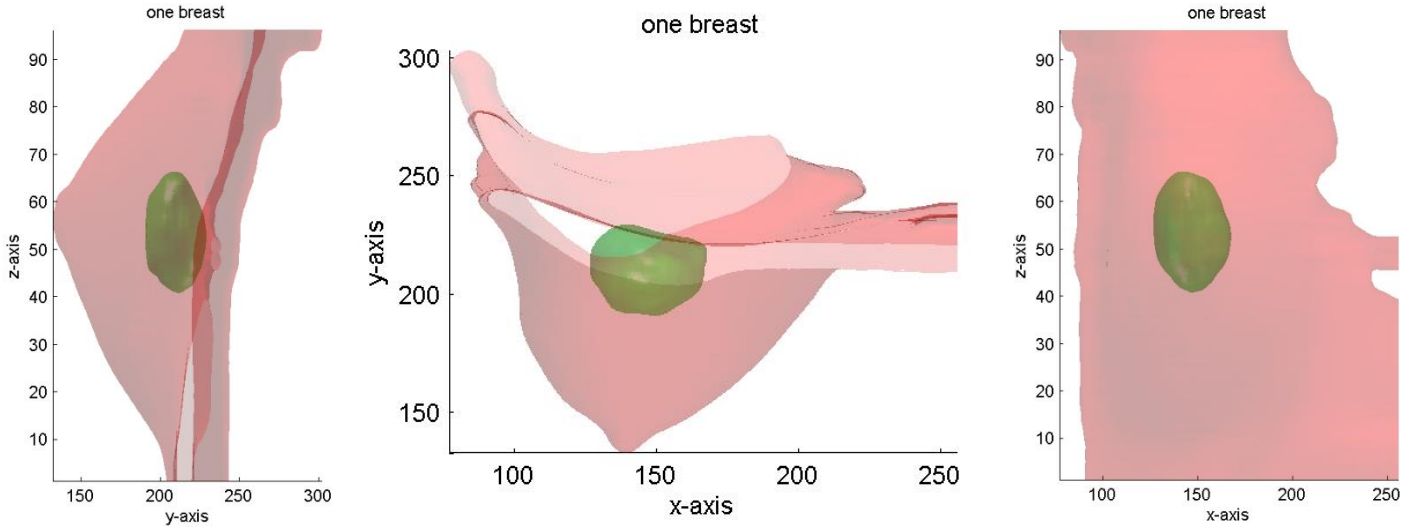


Figure 2.3.3 Three views in 3D model of a breast and the tumor, where pink area represents breast and green area represents tumor.

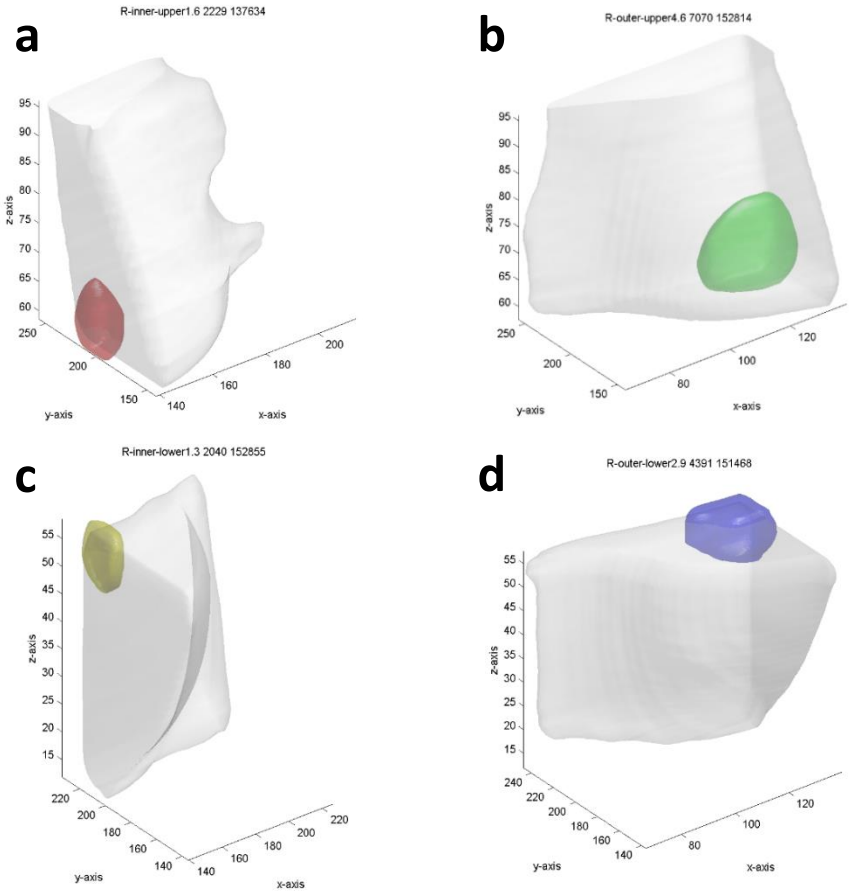


Figure 2.3.4 Quadrant segmentation result of the case in **Figure 2.3.3. (a), (b), (c), (d)** stands for the inner upper, outer upper, inner lower and outer lower quadrant of the breast, respectively.

2.4 Methodology of breast tumor segmentation

The lesions were segmented from contrast enhancement maps, obtained by subtracting pre-contrast images taken at the third frame from post-contrast images taken at the sixth frame. **Figure 2.4.1** shows the flowchart of the segmentation procedures for mass tumors and non-mass tumors, respectively.

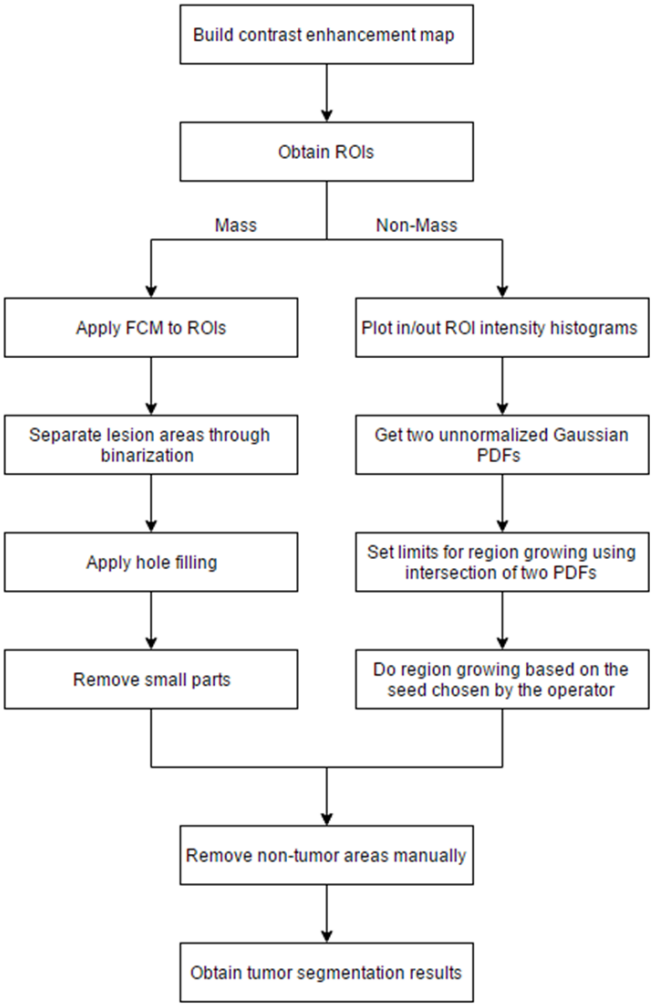


Figure 2.4.1 Flow chart of segmentation procedures for mass and non-mass tumors.

Firstly, the operator reviewed all images in an image sequence to determine the lesion location and the beginning/ending slices containing the tumor. Then, a rectangle box was manually placed over the lesion location as an initial region of interest (ROI). After this, different steps were taken for mass tumors and non-mass tumors.

2.4.1 Mass tumor segmentation

For mass tumors, fuzzy-C-means (FCM) clustering-based algorithm was applied. Developed by J.C. Dunn in 1973, FCM is one of the most used fuzzy clustering algorithms applied to feature analysis, clustering, and classifier designs without supervision. It classifies the image by attempting to partition a finite collection of elements in the feature space into clusters. The clustering is achieved under some criterion by iteratively minimizing a cost function that is dependent on the distance of the pixels to the cluster centroids. [2.7] The cost function is denoted as below:

FCM clustering

$$\mathrm{J} = \sum_{j=1}^{N}\sum_{i=1}^{c} u_{ij}^{m}\left\|x_j - v_i\right\|^2$$

Where u_{ij} represents the membership of pixel x_j in the ith cluster, v_i is the ith cluster center. In this equation, m is a constant that controls the fuzziness of the result and $\mathrm{m} = 2$ is usually used for FCM clustering.

The segmentations were taken in these steps:

(1) Enhance lesion area within the ROI using an un-sharp filter with a 3 by 3 kernel constructed using the inverse of two-dimensional Laplacian filter;

(2) Apply FCM on the enhanced ROI to obtain membership map of all voxel, which indicates the probability of each voxel belonging the lesion cluster or the non-lesion cluster. The weighting component on each fuzzy membership is chosen as 2 and the stopping criteria was set as the absolute change in objective function in consecutive iterations less than a pre-specified limit $\varepsilon = 10^{-5}$;

(3) Binarize the lesion membership map with the selected threshold to differentiate between lesion and non-lesion voxels;

(4) Remove scattered voxels that are not connected to the main body of lesion and apply hole filling to include the voxels within the main body.

2.4.2 Non-mass tumor segmentation

For non-mass lesions, FCM did not work well because of the scattered boundaries and irregular shape of the lesion. Therefore, additional procedures need to be applied:

(1) Draw out and compare the signal intensity histograms of fibroglandular tissues within and outside the rectangle ROI;

(2) Fit two unnormalized Gaussian probability density functions (PDF) [2.8] to normalize the difference in the pixel numbers of two histograms;

(3) Operator manually selected a seed inside the rectangle box for region growing, with the threshold determined by the intersection of two Gaussian probability density functions;

(4) Remove scattered voxels that are not connected to the main body of lesion and apply hole filling to include the voxels within the main body. **Figure 2.4.2** illustrates the procedures for non-mass tumor segmentation.

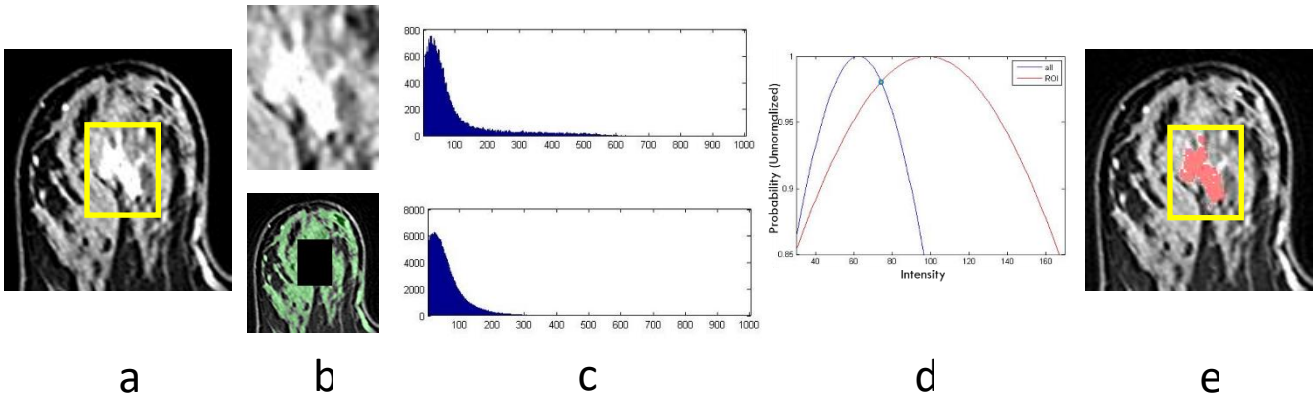


Figure 2.4.2 Procedures for non-mass tumor segmentation. **(a)** Contrast-enhancement map with manually placed tumor rectangle box. **(b)** The top picture shows pixels inside the rectangle box, and the bottom picture shows fibroglandular tissues outside the box, highlighted in green. **(c)** The histograms obtained from pixels inside and outside the box from corresponding pictures in (b). **(d)** Unnormalized PDF [2] is fitted in order to normalize the difference in the pixel numbers of two histograms, and the intersection of two PDF is set as the limit of region growing. **(e)** Red area shows the final segmented result of this non-mass tumor.

After the automated segmentation steps, the outline of the lesion ROI on each slices between pre-defined beginning/ending slices were displayed in the lower window. If necessary, the operator can choose to manually perform corrections using the modification

dialogue, including inclusion and exclusion. **Figure 2.4.3** shows the main interface of the segmentation program in this study.

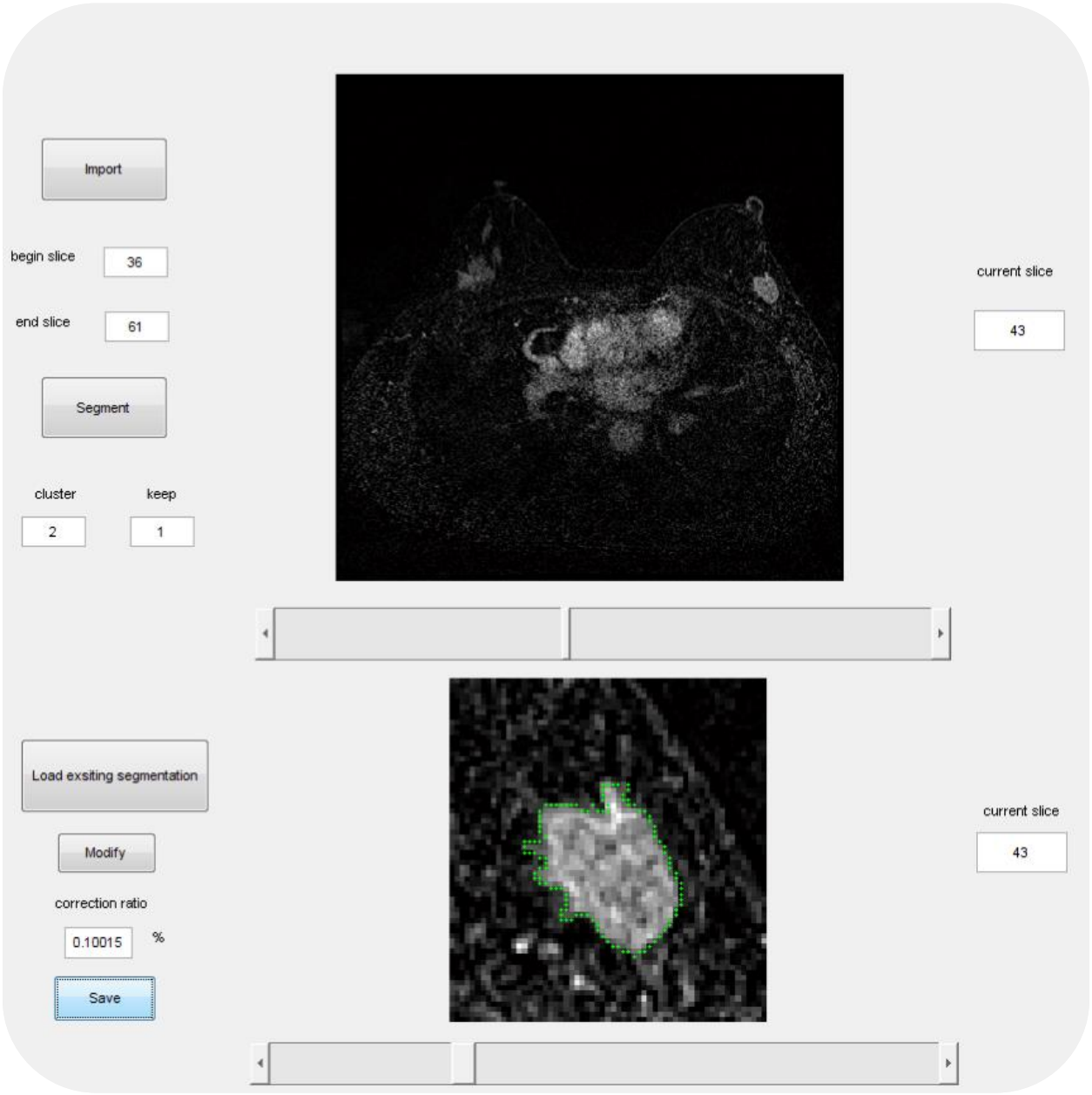


Figure 2.4.3 Main interface of the segmentation program. 'Import' button is used for import image sequence and the operator needs to decide the begin/end slice of tumor after reviewing the whole image sequence. After inputting the FCM clusters, the results of segmentation will display in the lower window. 'Modify' button can be used for either inclusion or exclusion manually by operator. After all, the correction ratio will be calculated and shown in the left lower corner.

2.5 Results

The segmentation results for 4 mass tumor cases are shown in **Figure 2.5.1**. **Figure 2.5.2** shows segmentation results of two non-mass tumors. **Figure 2.5.3** shows a 3D model of the segmented non-mass tumor.

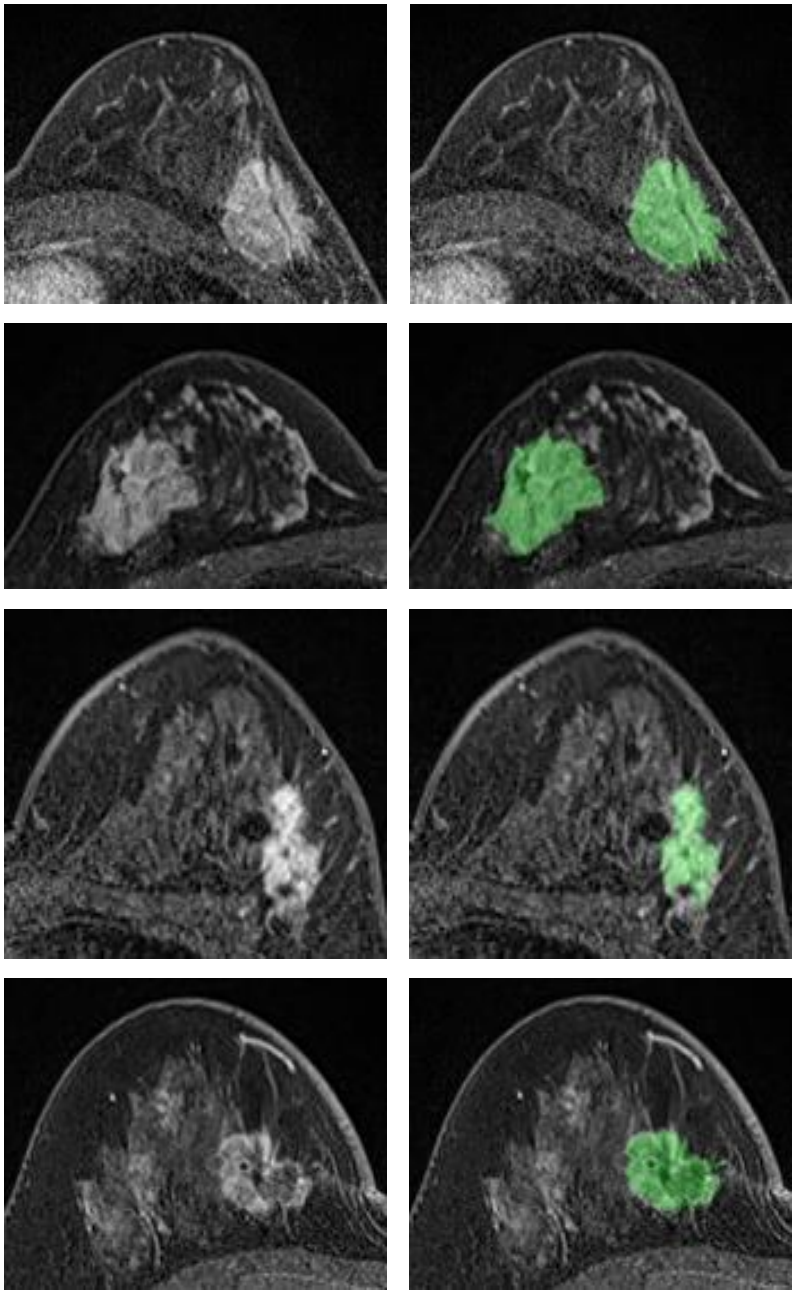


Figure 2.5.1 Example of four mass tumor segmentation results. Left column are the original post-contrast images. Green areas in right column show the segmented mass lesions.

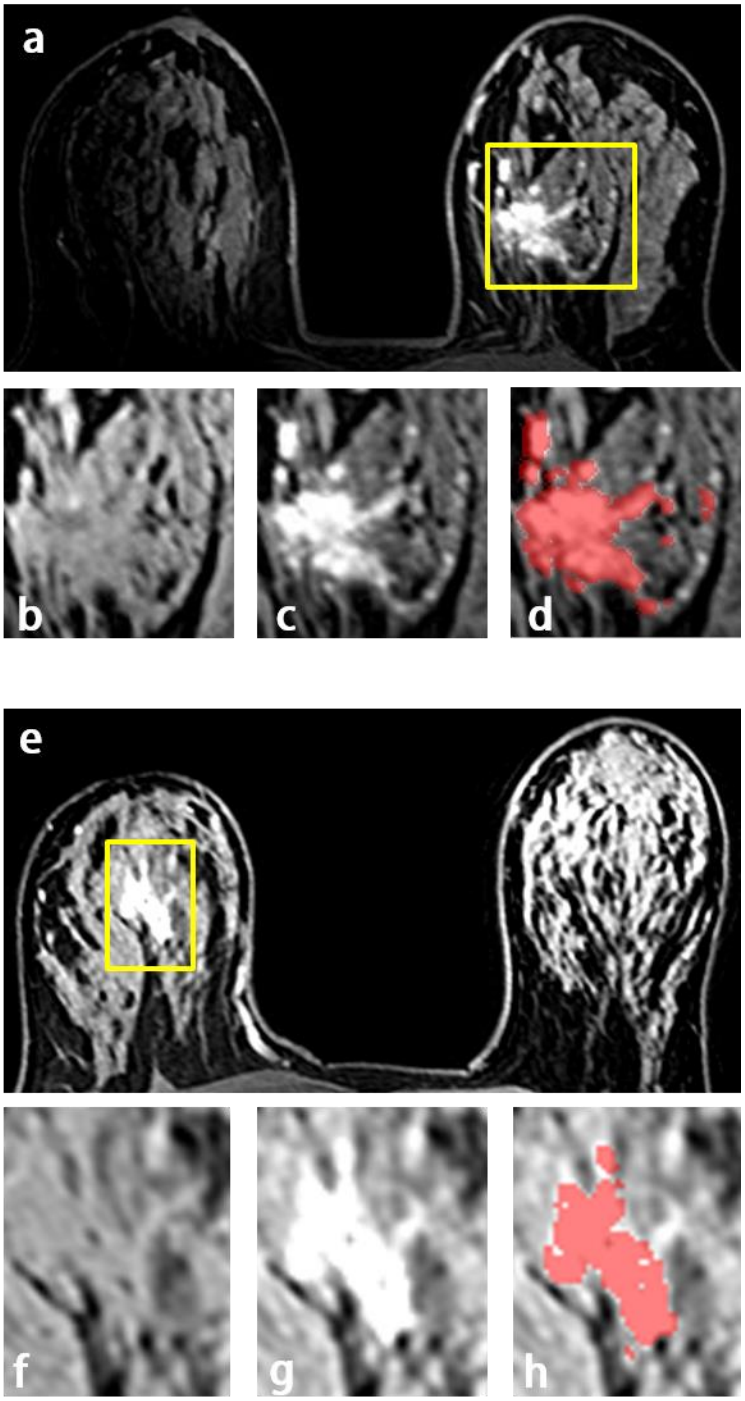


Figure 2.5.2 Example of two non-mass tumor segmentation results. (a)(e) Post-contrast images with manually placed tumor rectangle box. (b)(f) Pre-contrast images of the tumor box. (c)(g) Post-contrast images of the tumor box. (d)(f) Red area shows the segmented non-mass lesions.

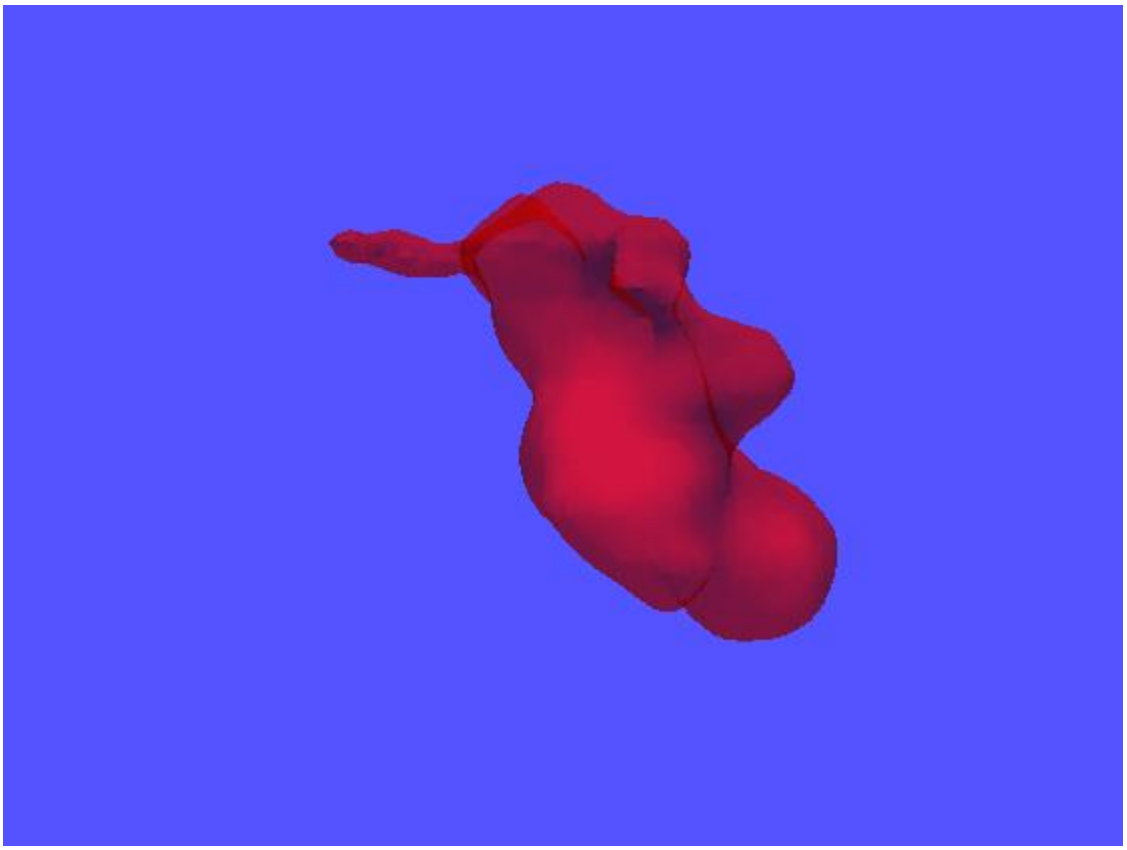
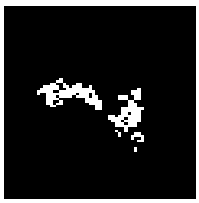
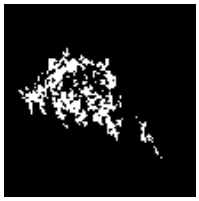


Figure 2.5.3 3D model of segmented non-mass tumor (the movie is included in the appendix of the thesis)

The segmentation quality was reviewed by an experienced radiologist and found to be satisfactory. Manual correction was rarely needed (less than 20% of the total cases; and if correction was needed the corrected pixels was fewer than 5% of total tumor pixels). 16 examples of the comparison between pre-edited and post-edited lesion areas are shown in **Table 2.5.1**.

Table 2.5.1 Examples of comparison between pre and post edited lesion areas

Case Number	Correction Rate	Original tumor mask	Edited tumor mask
1	6.299213		
2	16.860465		
3	2.378256		
4	0.761630		
5	0.229621		

6	0.123286		
7	0.384455		
8	3.283788		
9	0		N/A
10	0.908575		
11	0.376277		

12	2.905875		
13	2.406089		
14	1.465690		
15	18.242123		
16	0		N/A

2.6 Summary and Discussion

We developed a new method for segmentation of non-mass lesions on breast MRI, based on region-growing with the threshold determined by comparison of the intensity histograms in an ROI containing tumor vs. outside ROI containing normal fibroglandular tissue. The fibroglandular tissue segmentation was performed by using an automatic template-based method. Of 122 cases, only 24 cases needed to be segmented using the non-mass segmentation method and the tumor segmentation quality was satisfactory as the correction rate is 3.54. However, to test the stability and accuracy of this algorithm, more non-mass-like cases should be included and improvements could be made when problems come up. The major problem of false-positive pixels in existing cases came from vessels and breast boundaries that displayed bright signal similar to that of tumors. Even though for those cases that needed manual corrections, fewer than 5% pixels compared to the total number of segmented tumor pixels were corrected, other studies, like finding vessels using neural network [2.9], should be included to decrease the false positive rate in this study.

Chapter 3 Spinal cord cancer segmentation

3.1 Introduction

In this study, there are seven types of spinal cancer being studied, including chordoma, giant cell tumor (GCT), granuloma, lymphoma, metastasis cancer, myeloma, spinal tuberculosis (TB).

Chordomas are generally specified as slowly-grown, localized rod-shaped, cartilage-like neoplasms arising from cellular remnants of the notochord. Although chordomas are relatively uncommon in spinal cancers, they are comparably radioresistant, require high doses of radiation to be controlled and tend to recur after treatment. [3.1]

Characterized by the presence of multinucleated giant cells, most giant cell tumors are a kind of benign tumors which usually grow from the epiphysis of long bones, typically femur and tibia. Giant cell tumors are usually diagnosed based on biopsy findings and their common treatment is curettage. [3.2]

Due to an abundance of plasma cells, granuloma is also known as plasma cell granuloma. It is a small area of inflammation in tissue appearing in my diseases, which is often the result of an infection and frequently occur in lungs. Whether or not granulomas contain necrosis is a crucial feature, since those containing necrosis tend to have infectious causes. [3.3]

Lymphoma is a systemic disease, more specifically, a cancer of immune cells called lymphocytes. Two main categories of lymphoma are Hodgkin's lymphomas (HL) and the non-Hodgkin lymphomas (NHL). Spinal lymphoma may have different growth patterns

like the involvement of paravertebral lymph nodes, malignant infiltration of the vertebral bone marrow and probable hematogeneous involvement of the epidural space. [3.4]

Being the third most common site for cancer cells to metastasize only after the lung and the liver, spine is a common place where spinal metastasis can be found and cancers are diagnosed. Bringing pain, fracture, neurological deficits and mechanical instability, spinal metastasis is one of the largest leading causes of morbidity in cancer patients, giving only about 10 months for patients to survive. [3.5]

Myeloma is also called multiple myeloma or myelomatosis. It is a cancer of plasma cells in the bone marrow. Such cancerous cells make body unable to produce enough normal blood cells and thus lead to anaemia, infections and so on. With common treatment like chemotherapy, symptoms can be eased and patients' survival time can be prolonged for a few years. [3.6]

With higher probability to be found in children and young adults, spinal tuberculosis is a kind of extrapulmonary form of disease. In all the cases of musculoskeletal, spinal tuberculosis accounts for almost 50%. Although they occur outside the lungs, diseases are seen in the vertebrae and can cause destruction to the intervertebral disk space and adjacent vertebral bodies. [3.7, 3.8]

3.2 Subjects and MRI Protocols

Chordoma(13), Giant Cell Tumor(26), Granuloma(17), Lymphoma(7), Metastatic cancer(80), Myeloma(9), Tuberculosis(24) cases were studied in this project.

The MRI studies were performed using a Siemens 1.5T system. The imaging sequences included axial T1WI and axial T2WI were acquired using fast spin echo. Each of

the sequence contains 30 images while their voxel size are mostly 0.6*0.6*1.0 (mm3), but may vary from case to case.

The contrast agents, 0.1 [mmol/kg] Gd-DTPA, was injected after four set of pre-contrast images was acquired, by using an Ulrich power injector at a rate of 2 ml/s followed by 20 cc saline flush at the same rate. A total of 12 frames were acquired, so the total DCE-MRI acquisition time period ranged from 120 to 168 seconds. Depending on the abnormal area that needed to be covered and the choice of sequences to acquire different views, the total imaging time was approximately 25-30 minutes.

3.3 Methodology of spinal lesion segmentation

For spinal cancer cases, contrast agent can enhance the exact location of spinal tumor. However, it is extremely difficult to observe the location and the boundaries of lesions in axial view by using simple segmentation algorithm. In this study, a method based on registration, normalized cut and region growing was developed.

3.3.1 Registration from sagittal view to DCE axial sequence

Dynamic contrast enhancement (DCE) is a locally used technique by which images of the resultant changes in signal intensity can be analyzed [3.9]. In this case, an approximate area needs to be found under axial view while lesion locations will be found in sagittal view and then transferred to axial view. Firstly, MRI images under sagittal view are used to

determine the range of operation in vertical axis. Secondly, complete vertebrae boundaries cannot be seen under sagittal view, rendering it tricky to determine the locations as well as shapes of lesions. In addition, organs like the liver and the spleen might obscure the distinction of lesion area under axial view. In order to obtain the approximate location of the lesion, the operators need to draw a vertebra containing the lesion area. The segmentation results were saved as masks for further uses.

For each slice of sequence in sagittal view, the corresponding coordinate of each pixel can be obtained, after which a large 3D matrix containing the 3D corresponding coordinates of approximate lesion area can be generated. In order to calculate the absolute global coordinates of this volume, the absolute global coordinate of the pixel point on the top left of this image stack needs to be calculated using ImageJ. After acquiring the 3D global coordinates as discussed above, a 3D bounding box is computed to determine the minimum volume that contains the lesion. The same procedure can be easily adapted to obtain the coordinate of the top left pixel point in DCE axial sequence as well as a large 3D matrix of axial bounding box. The min distance function will then be applied to find the corresponding point in axial bounding box with minimum distance to the corresponding point of 3D sagittal tumor model. Due to the difference of thickness and resolution, the result might not be connected and ideal enough. For the final step, smoothing process were taken in order to obtain consecutive and smooth tumor masks in axial view.

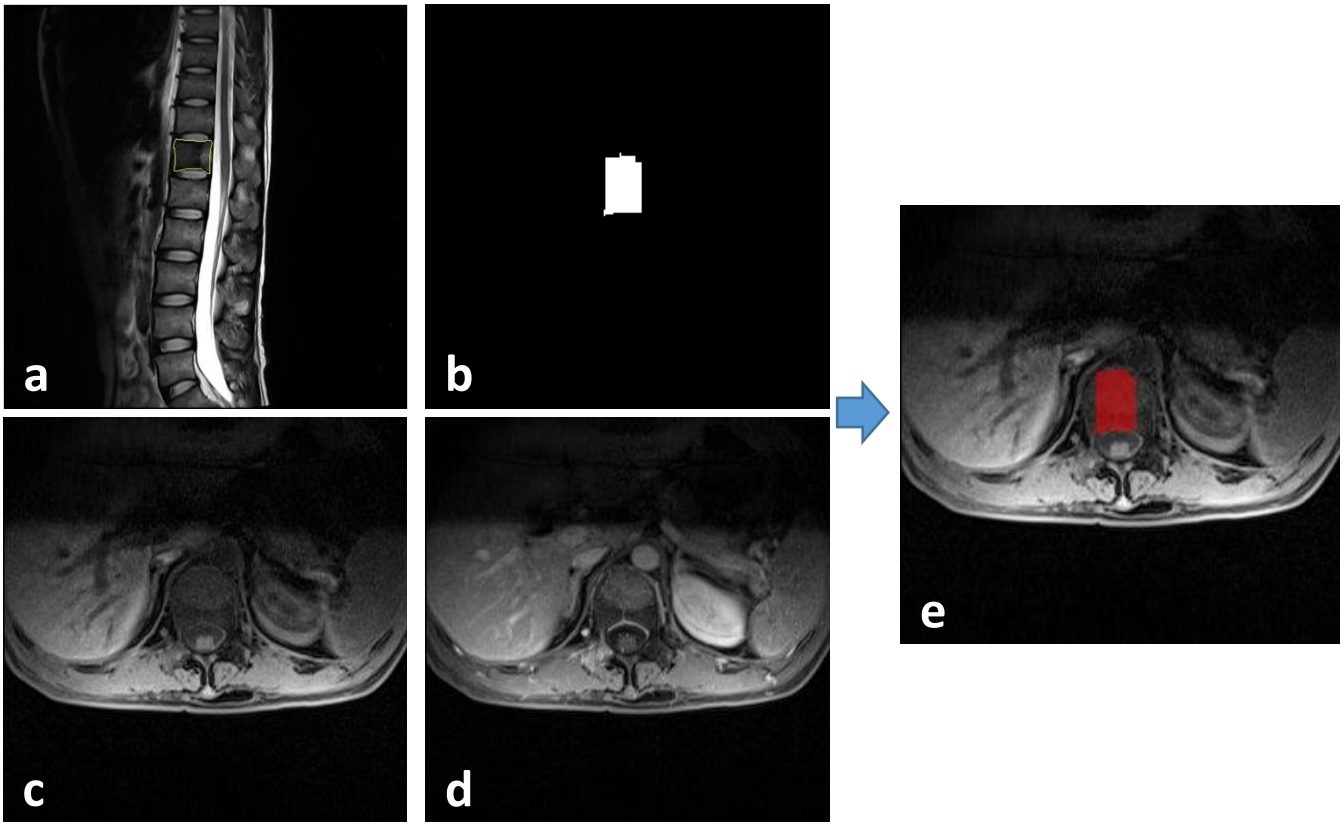


Figure 3.3.1 Basic steps showing registration from sagittal view to DCE axial sequence. (a) Original lymphoma sagittal view MRI with yellow manual-drawn ROI. (b) Registration result mask in DCE axial view. (c) Corresponding slice of the first frame DCE-MRI. (d) Corresponding slice of the fourth frame DCE-MRI. (e) Colorcoded picture using registration result mask and the fourth frame DCE-MRI.

3.3.2 Normalized Cut

Normalized cut is an unsupervised segmentation technique not requiring initialization developed by Shi and Malik. [3.10] In NCut, each voxel is considered a node. It is based on a global criterion and approaches the segmentation as a graph-partitioning problem. Instead of looking at the value of total edge weight connecting the two partitions, the cut cost is computed as a fraction of the total edge connections to all the nodes in the graph. Also, it maximizes both the total dissimilarity between the different groups and the total similarity within the groups. A measure of disassociation, also known as dissimilarity

between two nodes, can be found to create an edge. Such disassociation measure is defined as normalized cut. [3.11]

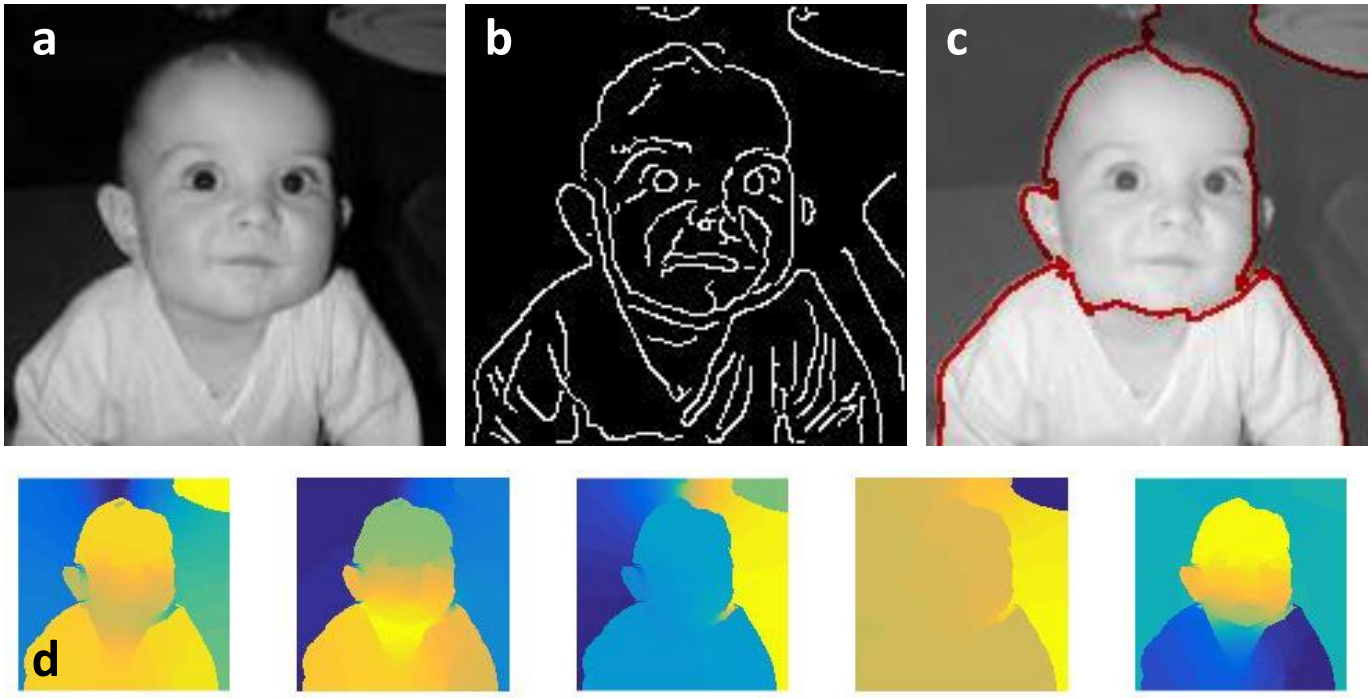


Figure 3.3.2 An example picture using normalized cut algorithm. (a) Original picture. (b) Edge detection result. (c) Segmentation result based on (b). (d) Eigenvectors of normalized cut of this picture.

NCut can also be expressed in formula below:

$$\mathrm{Ncut(A,B)} = \frac{cut(A,B)}{assoc(A,Z)} + \frac{cut(A,B)}{assoc(B,Z)}$$

where

$$\mathrm{assoc(A,Z)} = \sum_{u \in A, p \in Z} W(u,p)$$

and similar as assoc(B, Z).

Using the definition above, a box that must include the lesion area needed to be decided to apply normalized cut and following processing steps. The height of the box was

determined by the top slice and bottom slice of ROI in the coronal direction. The width of the box is determined as 5 times the distance of the overall thickness of slices containing ROIs in sagittal view. the farthest points to the middle in the sagittal direction. After deciding the calculation box, normalized cut was applied on every slice in sagittal view. Having the masks of registration result in axial view, an overlapped picture between axial lesion estimated mask and normalized cut result can easily be obtained. The procedure was shown in **Figure 3.3.3.**

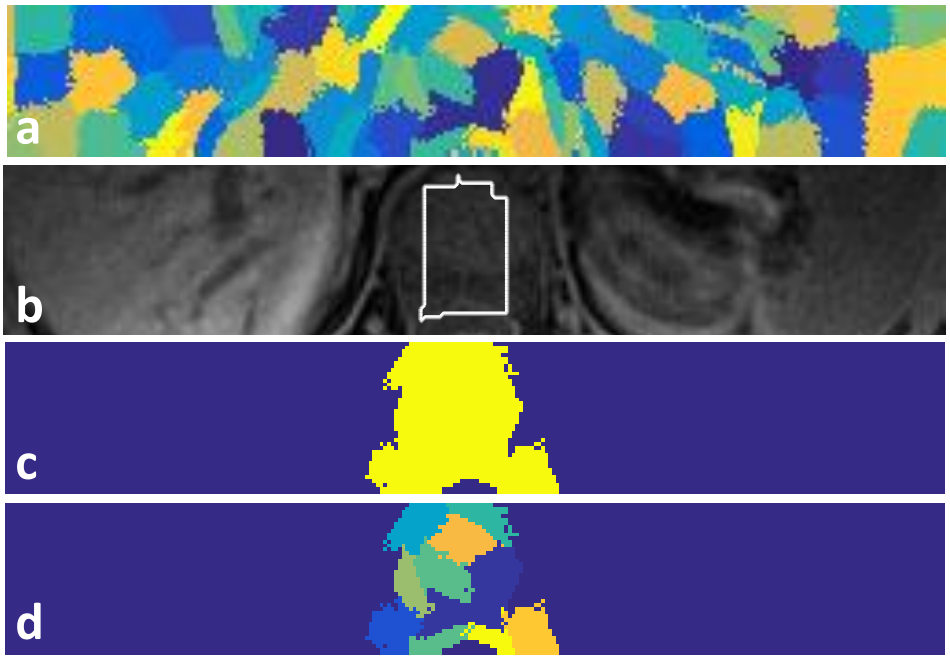


Figure 3.3.3 Lymphoma patient case using normalized cut algorithm. (a) Original picture after edge detection. Each color lump represent a different part in the image. (b) Registration result contured on original image. (c) Mask containing overlapped color lumps. (d) Color lumps overlapping with registration result.

3.3.3 Region growing

Seeded region growing approaches to perform segmentation of an image into regions with respect to a set of points, also known as seeds. Starting with a given set of seeds, S1, S2, . . ., Sq, SRG then in the later each step involves one additional pixel to one of the seed sets. It also finds a tessellation of the image into regions, R1, R2, . . ., Rq, where the regions are chosen as homogeneous as possible. All pixels in the same region are labelled as allocated labels, while the rest of pixels are called unallocated pixels. Let H be the set of all unallocated pixels which are adjacent to at least one of the labeled regions.

$$H = \left\{ (x,y) \notin \bigcup_{i=1}^{q} R_i | N(x,y) \cap \bigcup_{i=1}^{q} R_i \neq \emptyset \right\}$$

where N(x,y) is set of immediate neighbors of pixel (x,y).

If N(x,y) meets two or more of the labeled regions, φ(x,y) is assigned a value i such that N(x,y) meets Ri and φ (x,y,Ri) is minimized.

$$\varphi(x,y) = \min_{(x,y)\in H} \{ \varphi\ (x,y,R_i) | j \in \{1,\dots,q\} \}$$

Before all pixels in the image have been allocated to corresponding regions, the seeded region growing procedure will continue. After that, the image is divided into a set of regions on the basis of the given constraints. Although seeded region growing algorithm requires high computation as well as problems in seed selection and pixel sorting, it is robust and free of tuning, which turns it quite suitable for automatic image segmentation. [3.12, 3.13]

In this study, the brightest point was chosen as the seed of region growing. Based on the region growing principle, threshold = 7 were used for most of the cases. Removing

scattered voxels that are not connected to the main body of lesion was the last step before obtaining the final axial tumor mask. Figures of these steps were shown in **Figure 3.3.3.1**.

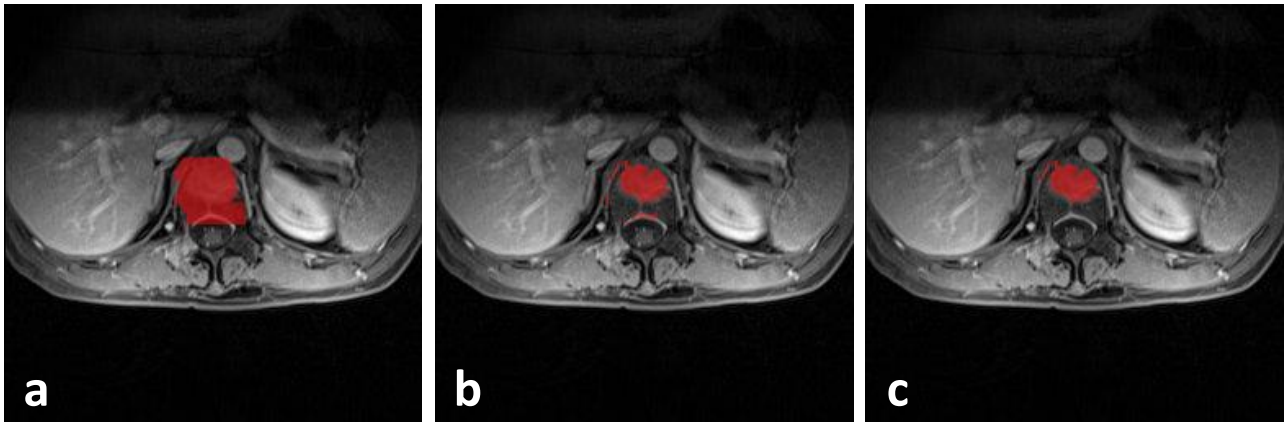


Figure 3.3.4 Lymphoma patient case using region growing algorithm. (a) Original picture with normalized cut result (red). (b) Red part represents region growing result. (c) Red part represents (b) after removing scattered boundaries.

3.4 Results

In this part, the method was applied to seven different spinal cases. **Figure 3.4.1** - **Figure 3.4.7** are the results where green areas represent the estimated vertebrae area and red areas stand for lesion area.

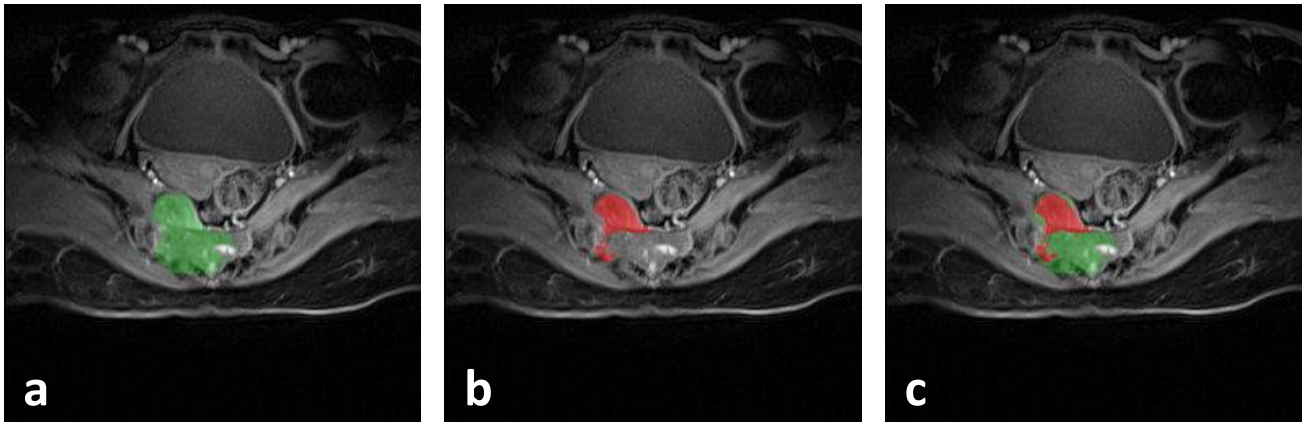


Figure 3.4.1 Chordoma Case.

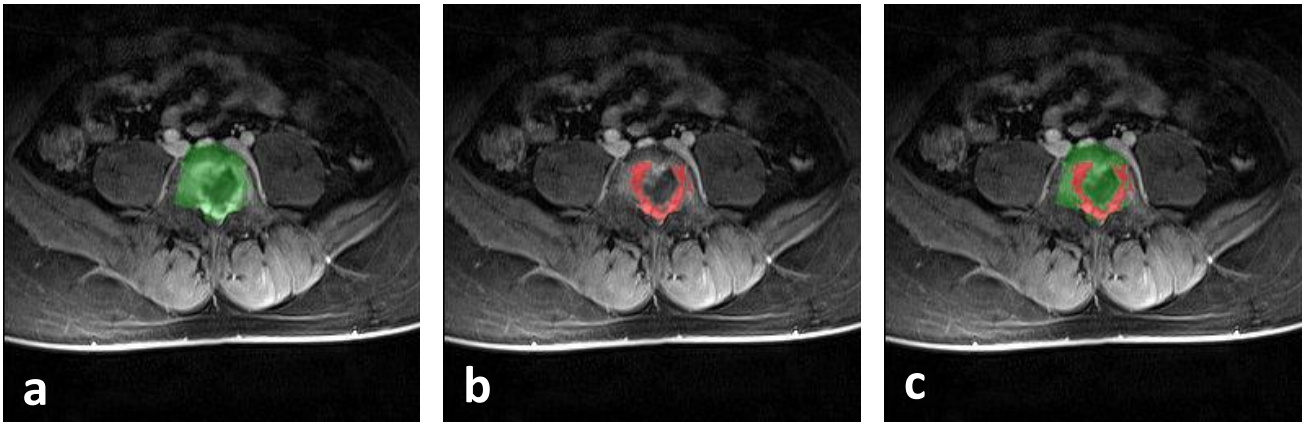


Figure 3.4.2 Giant Cell Tumor (GCT) Case.

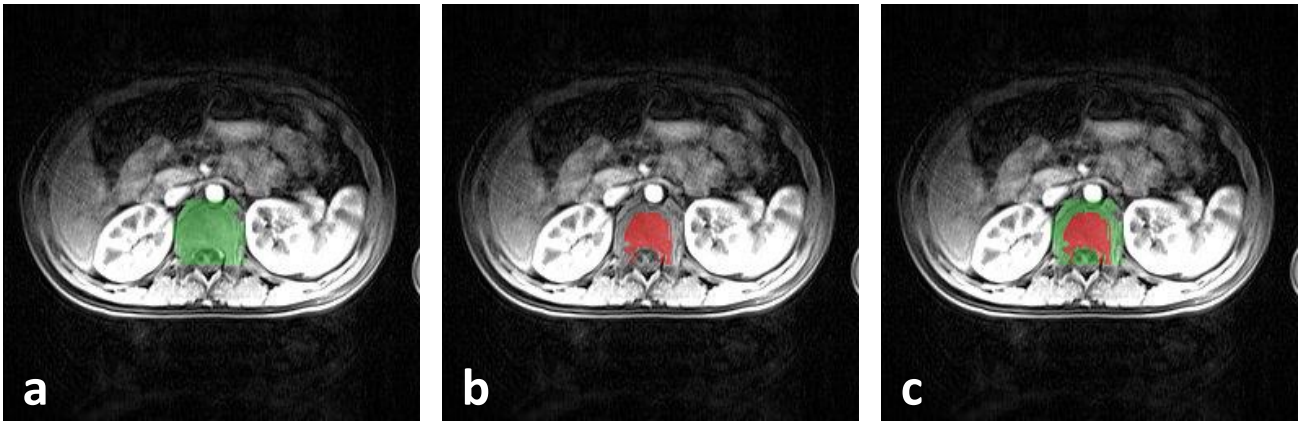


Figure 3.4.3 Granuloma Case.

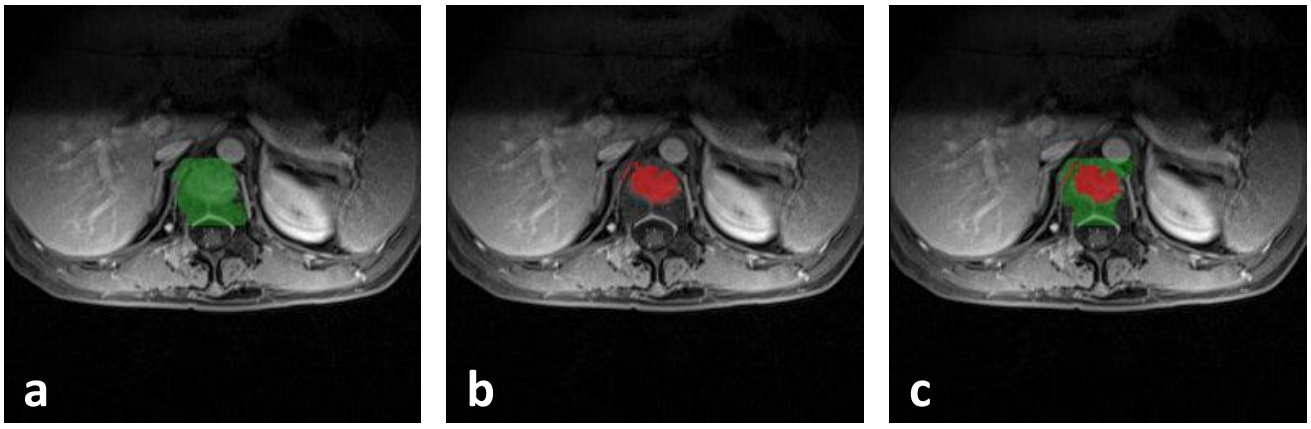


Figure 3.4.4 Lymphoma Case.

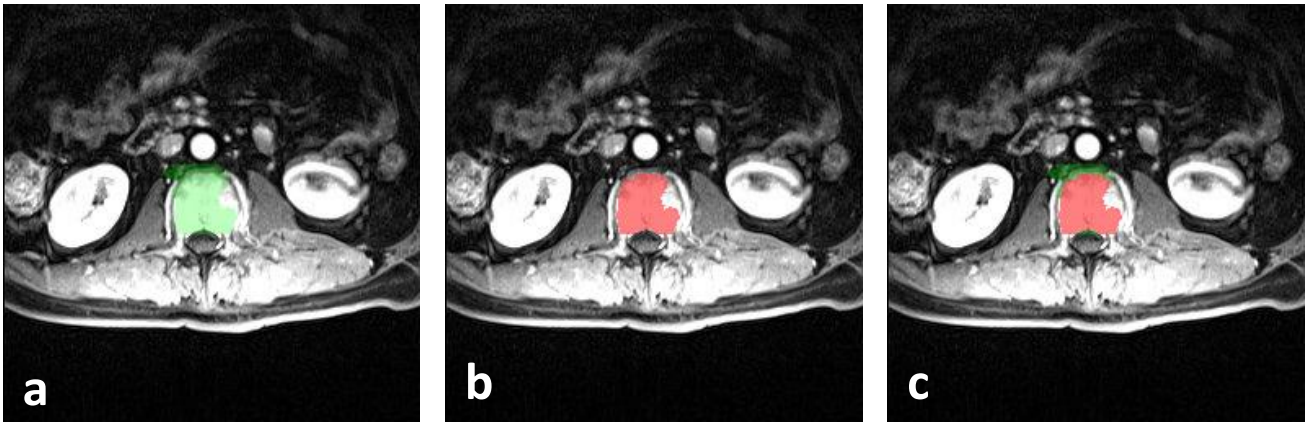


Figure 3.4.5 Metastatic cancer Case.

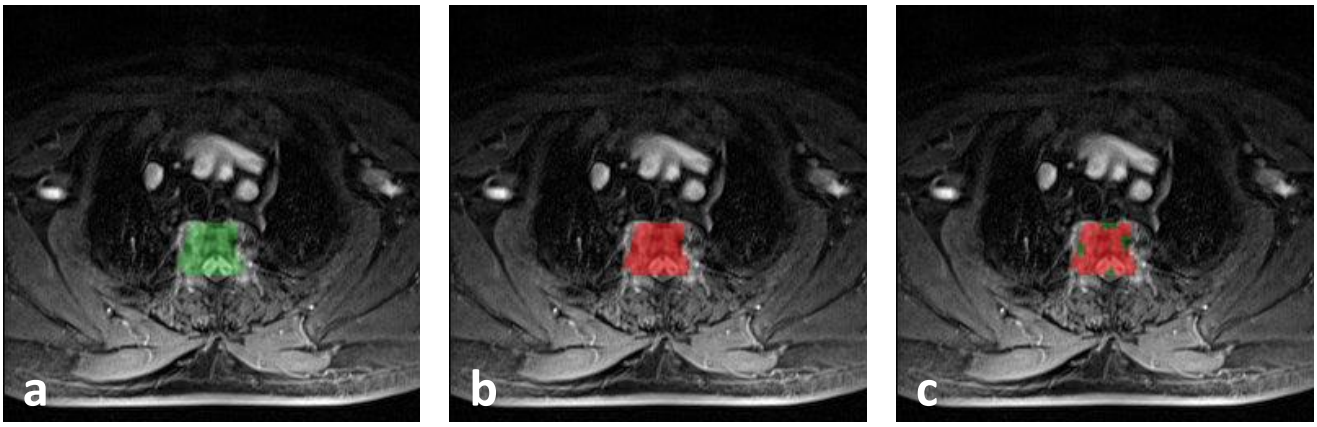


Figure 3.4.6 Myeloma Case.

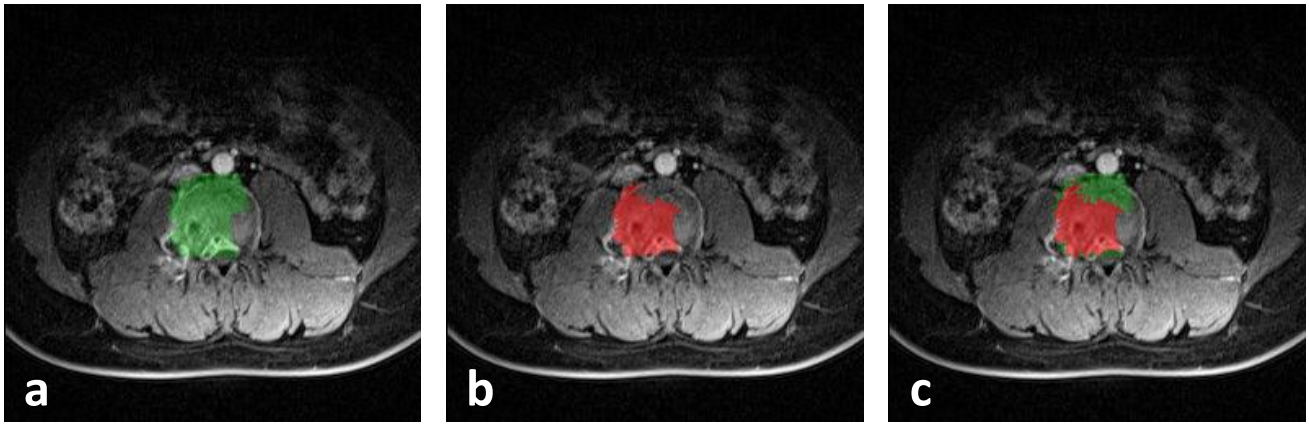


Figure 3.4.7 TB Case.

3.5 Summary and Discussion

In this study, seven types of spinal cancer were included. Firstly, one experienced radiologist drew the estimated vertebrae body containing tumor area manually in sagittal view. Secondly, registration from sagittal view to axial view was applied using 3d bounding box corresponding theory. Then normalized cut could segment the picture in different parts using the picture's characteristics. Finally, region growing technique used the brightest point as the seed and a stable threshold to grow out the lesion region.

In this study, loss of cases in this study could be due to the DCE sequence were taken in the wrong part of vertebra, which means that there were no registration results in the first step. Another problem occurred was that registration process was extremely time-consuming, especially for those cases that the tumors were very large. For some cases, it even took more than 15 hours to compute the results in axial view. This problem should be improved in future study by using a different algorithm or raising a more time-efficient method to approach this problem. Moreover, for most of the cases, the final results were claimed satisfactory by an experienced radiologist. However, in some cases, due to the body movement between the pre-contrast frame and post-contrast frame, the enhancement maps were so blurry that the region growing part could not find the correct brightest point as the seed. Furthermore, it could also lead to growing too much or too less for the tumor area because of the unclear boundaries.

Chapter 4 Machine Learning

4.1 Introduction

Machine learning is the study of computational methods in order to improve performance by mechanizing the acquisition of knowledge from experience. Using algorithms that iteratively learn from data, machine learning gives computers the ability to find hidden insights without being explicitly programmed where to look. Aiming at providing increasing levels of automation while replacing much time-consuming human labor with techniques which can improve the accuracy and efficiency in training data, machine learning makes it possible to quickly and automatically produce models that can analyze bigger, more complex data and deliver faster, more accurate results. Five basic machine learning paradigms include neural network associated paradigm, instance-based or case-based learning, generic algorithms, rule induction and analytic learning. Besides, due to the difference in what learning "signal" is or what "feedback" is available to the learning system, machine learning tasks can be typically categorized into three types, namely supervised learning, unsupervised learning and reinforcement learning. [4.1, 4.2]

In this study, four texture analysis matrices (GLCM, GLRLM, GLSZM, NGTDM) were used to calculate the texture features within the 3D lesion ROI.

The **Gray Level Co-Occurrence Matrix Occurrence Matrix (GLCM)** characterizes the texture of an image by calculating how often pairs of pixel with gray-level (grayscale intensity or level) value *i* occurs either horizontally, vertically, or diagonally to adjacent pixels with the value with the value *j*. The **Gray Level Run Length Matrix (GLRLM)** is a method of extracting higher order statistical texture features. Rather than looking at pairs of

pixels, the GLRLM looks at runs of pixels. That is, how many pixels of a given grey value occur in a sequence in a given direction. The rows of the GLRLM represent grey levels in the image. However, the columns now represent lengths of runs, with the entries corresponding to the number of runs of the given length in the image. The **Grey Level Size Zone Matrix (GLSZM)** looks at zones of 9-connected pixels. That is, how many pixels of a given grey value are connected in a single group. The rows of the GLSZM represent grey levels in the image, and the columns now represent sizes. Thus the entries correspond to the number of groups of pixels of a given grey value that are a given size. **Neighborhood Gray Tone Difference Matrix (NGTDM)** defines features related to human perception of a texture and is mostly used to calculate higher-order parameters and describe the local features. It is based on the differences between each voxel and the neighboring voxels in the adjacent image planes. A NGTDM, s(i), is a column matrix formed by summing the absolute value of the pixel being observed minus the average of the pixels in its neighborhood. The neighborhood was predefined as a distance of d = 2 pixels. Once the NGTDM was formed and image intensities are resampled, five texture features were calculated, namely, the coarseness, contrast, busyness, complexity, and texture strength. [4.3, 4.4]

The random forest algorithm was utilized to compute the significance of each feature, and those with highest score were used for pair-wise classification by using a logistic regression model. In 1996, Breiman proposed bagging method to generate classifiers and aggregate their results. Later in 2001, he added an additional layer of randomness to bagging and called it random forest. Besides constructing each tree using a different bootstrap sample of the data, random forests changed the way that the classification or regression trees are constructed. Instead of splitting each node using the best split among all variables in

standard tress, each node is split in random forest using the best among a subset of predictors randomly chosen at that node. Also, what makes random forest very user-friendly is that it has only two parameters (the number of variables in the random subset at each node and the number of trees in the forest), and it is usually not very sensitive to their values. [4.5]

4.2 Breast cancer machine learning

After segmentations were done in chapter 2, the 3D ROI of the tumor was analyzed to obtain 11 morphology features (Volume, Surface Area, Compactness, Sphericity, NRL Entropy, NRL Ratio, Roughness, 3D Circularity, 3D Complexity, 3D Irregularity, 3D Compactness) and 53 texture features using GLCM, GLRLM, GLSZM, NGTDM metrics.

Histologically, breast cancer can be broadly categorized into in situ carcinoma and invasive (infiltrating) carcinoma. Invasive carcinomas are a heterogeneous group of tumors differentiated into several subtypes. The major invasive tumor types include infiltrating ductal, invasive lobular, ductal/lobular, mucinous (colloid), tubular, medullary and papillary carcinomas. Another classification of breast cancer is based on molecular subtypes, including Luminal A, Luminal B, Triple negative/basal-like, and HER2 type. [4.6] In this study, of all 122 cases, 94 had complete molecular biomarkers to be classified as triple negative (14 cases), HER2-positive (29 cases), Hormonal-positive and HER2-negative (51 cases).

In this chapter, machine learning was used to classify these three types of tumor based on the hidden information, so called features of the tumors in the images, for example

volume, surface area and grey level. Besides complicated texture features, the first thing should be considered was morphology features. [4.7] One-way ANOVA and random forest were respectively applied to morphological features analysis and texture analysis in order to find which features should be used in machine learning.

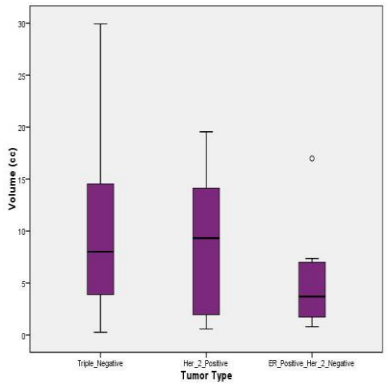
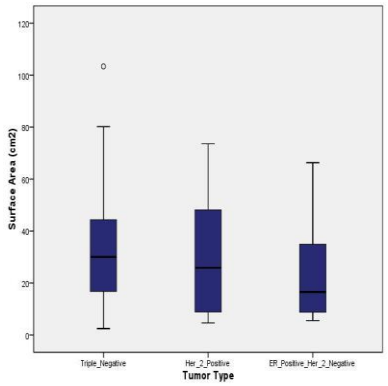
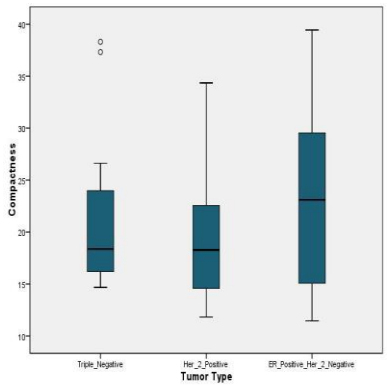
The one-way analysis of variance (ANOVA) is used to determine whether there exists any statistically significant differences between the means of two or more independent groups. Its null hypothesis lists as below:

$$H_{null}: m_1 = m_2 = m_3 = \cdots = m_n$$

where m represents the group mean and n stands for the number of groups. If there are at least two group means that are statistically significantly different from each other, the alternative hypothesis H will be accepted and it means the results of one-way ANOVA is statistically significant.

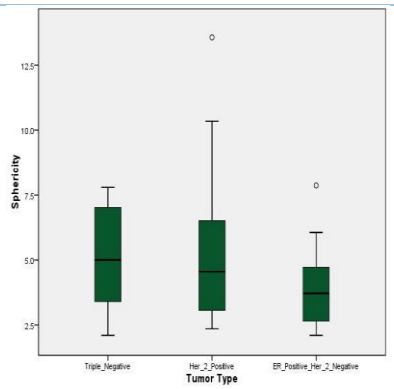
In the chart below, the definition of 11 morphological features and results of ANOVA analysis were included.

Table 4.2.1 ANOVA of 11 morphology features

Formula/Definition	Anova Result
Volume: $$Vol = \sum_{\nabla x,\nabla y,\nabla z} F_{ROI}(x,y,z) \cdot p^2 \cdot t$$	
Surface Area: $$Surf = \sum_{\nabla x,\nabla y,\nabla z} S_{ROI}(x,y,z) \cdot p \cdot t$$	
Compactness: $$Comp = \frac{Surf^{3/2}}{Vol}$$	

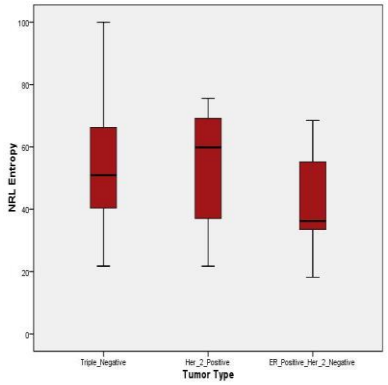
Sphericity:

$$Spher = \frac{\mu_{NRL}}{\sigma_{NRL}} \quad \mu_{NRL} = \frac{1}{N}\sum_{j=1\ldots N} r_j$$



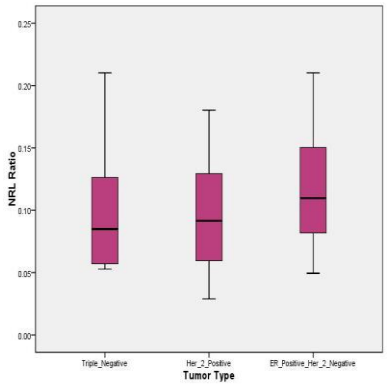
NRL Entropy:

$$E_{NRL} = -\sum_{j=\cdots H} prob_j log_2(prob_j)$$



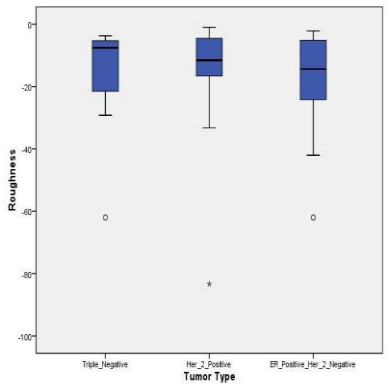
NRL Ratio:

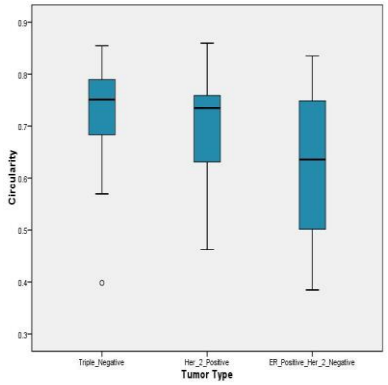
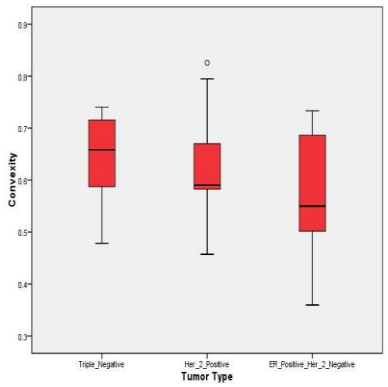
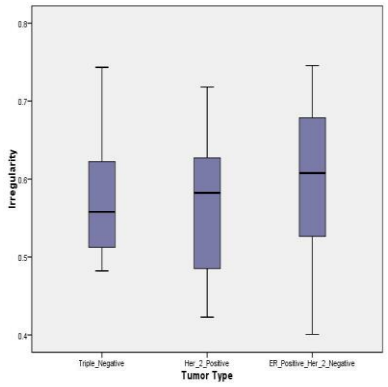
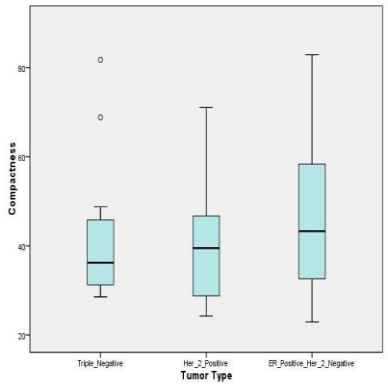
$$R_{NRL} = \frac{1}{N \cdot \mu_{NRL}} \sum_{i=1\ldots N} (r_j - \mu_{NRL})\, r_j > \mu_{NRL}$$



Roughness

$$= \frac{\sqrt[4]{\frac{1}{N}\sum_{j=1\ldots N}(r_j - \mu_{NRL})} - \sqrt{\frac{1}{N}\sum_{j=1\ldots N}(r_j - \mu_{NRL})}}{\mu_{NRL}}$$



Use the volume to calculate an equivalent sphere. **Circularity** is defined as the fraction of volume inside.	
Use gift-wrapping algorithm to obtain a minimum convex hull. **Convexity** = The fraction of volume inside the hull.	
Irregularity = 1- [(surface area of equivalent sphere) / (total surface area)]	
Compactness = Volume^(2/3)/area	

Out of 11 morphology features, eight features' p-value were lower than 0.05, which means they were significant enough to differentiate between three subtypes. These eight features combined with 50 texture features generated by using GLCM, GLRLM, GLSZM, NGTDM metrics were chosen to apply in machine learning. Based on cross validation method, ROC curves and AUC of pairwise comparison were generated. However, none of the single feature or combined features had the ability to differentiate among these three tumor subtypes. The area under the ROC curve (AUC) were shown in **Figure 4.2.1**.

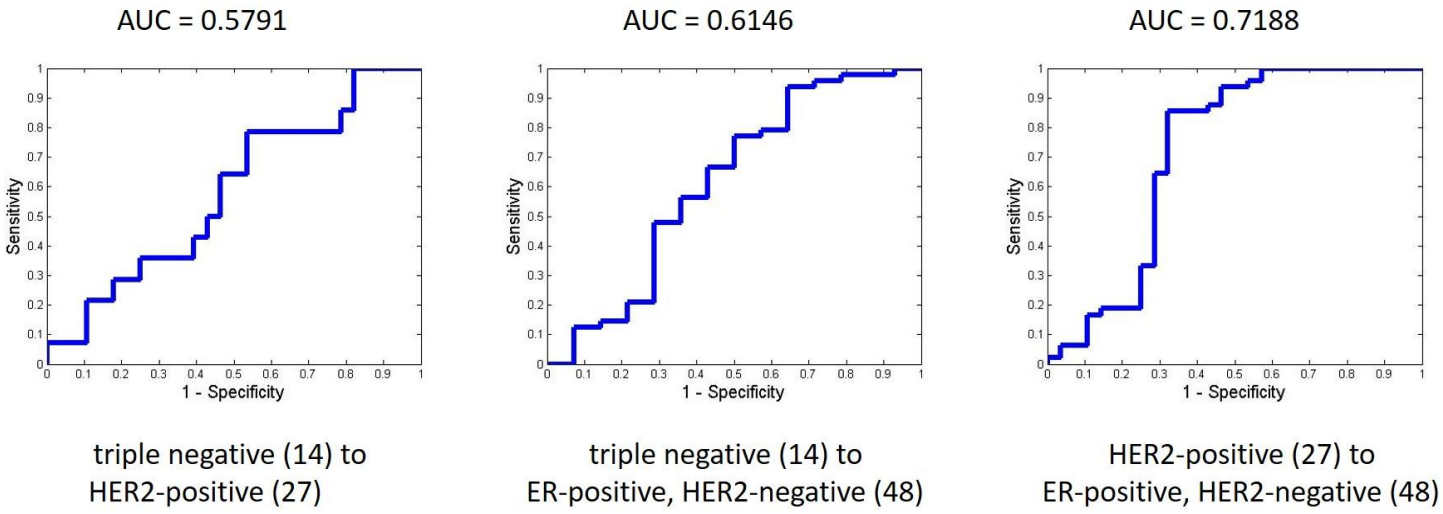


Figure 4.2.1 ROC curve and AUC among three tumor subtypes.

4.3 Spinal cancer machine learning

As mentioned in Chapter 3, there are seven types of spinal cancer dataset, including Chordoma(13), Giant Cell Tumor(26), Granuloma(17), Lymphoma(7), Metastatic cancer(80), Myeloma(9), Tuberculosis(24). Metastatic cancer cases were rejected in this study due to the

exceptionally high variability of lesion characteristics in MRI images. In order to avoid overfitting situation during training the classification model, GCT, Granuloma and TB were chosen as the datasets to be applied machine learning. After avoiding cases with insufficient information, GCT(15, 8 male, 7 female, mean age 34)), Granuloma(12, 6 male, 6 female, mean age 22) and TB(12, 6 male, 6 female, mean age 43) were used in this study.

Three normalized maps were used in this study:

- The maximum signal enhancement (SE) ratio normalized to the pre-contrast intensity [(Smax-S0)/S0] shown in **Figure4.3.1 (a)**.
- The steepest wash-in enhancement ratio [(S2-S1)/S0] (S1 and S2 were two adjacent time points that showed the greatest signal enhancement) shown in **Figure4.3.1 (b)**.
- The wash-out slope [(Slast-Speak)/Speak x 100%],if no peak using the signal intensity at 67 seconds as the reference [(Slast-S67s)/S67s x 100%] shown in **Figure4.3.1 (c)**.

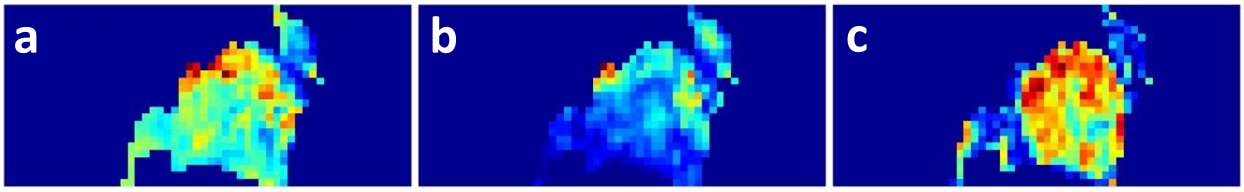


Figure 4.3.1 Examples of three maps corresponding to three definitions.

For each map, four texture analysis matrices (GLCM, GLRLM, GLSZM, NGTDM) were applied to calculate the texture features within the 3D lesion ROI. 50 texture parameters

(Shown in **Figure 4.3.2**) together with expectation, variance and entropy were extracted from these matrices in each map, so all together there were 159 features.

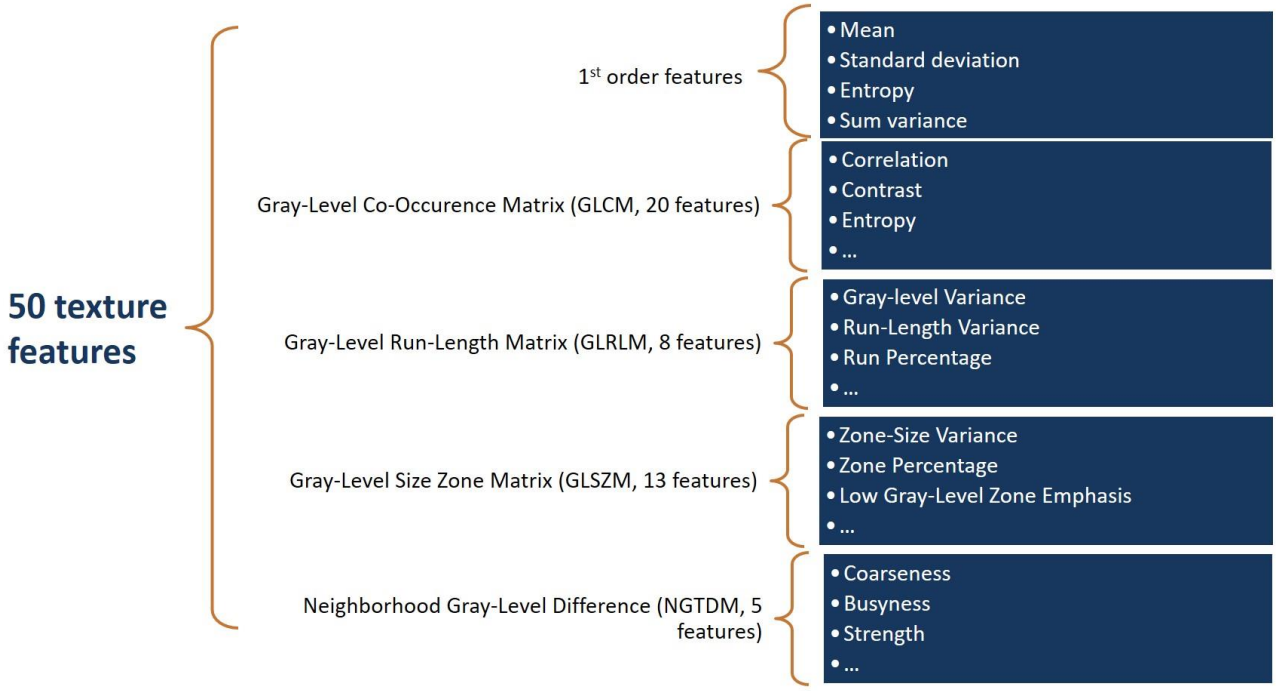


Figure 4.3.2 Brief introduction of 50 texture features.

Random Forest algorithm was then applied to choose 10 of most significant features that can differentiate among three kinds of spinal cancer. In this study, creation of bags of decision tress was first applied. Every tree in the ensemble is grown on an independently drawn bootstrap replica of input data. Plotting out the relationship between classification error and number of number of grown trees can lead to the optimal tree number corresponding to the minimum error. **Figure 4.3.2 – Figure 4.3.4** are the results of error-tree plots and the predictors' importance estimates.

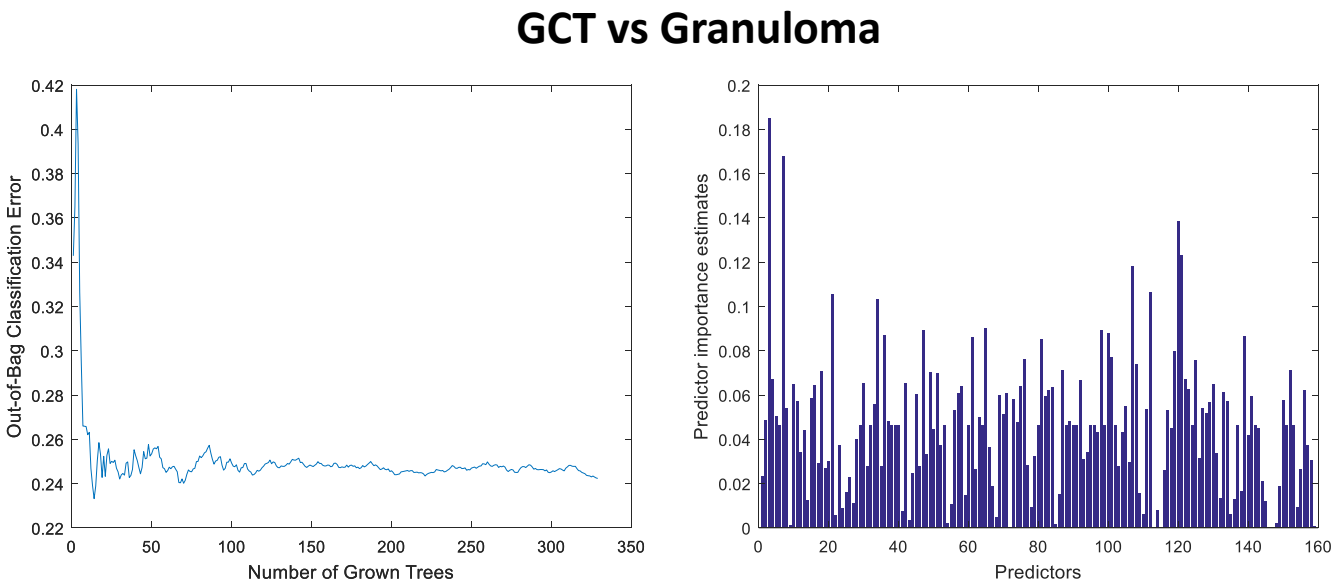


Figure 4.3.3 Classification Error and predictors' importance estimates between GCT and Granuloma. Tree number was set as 329.

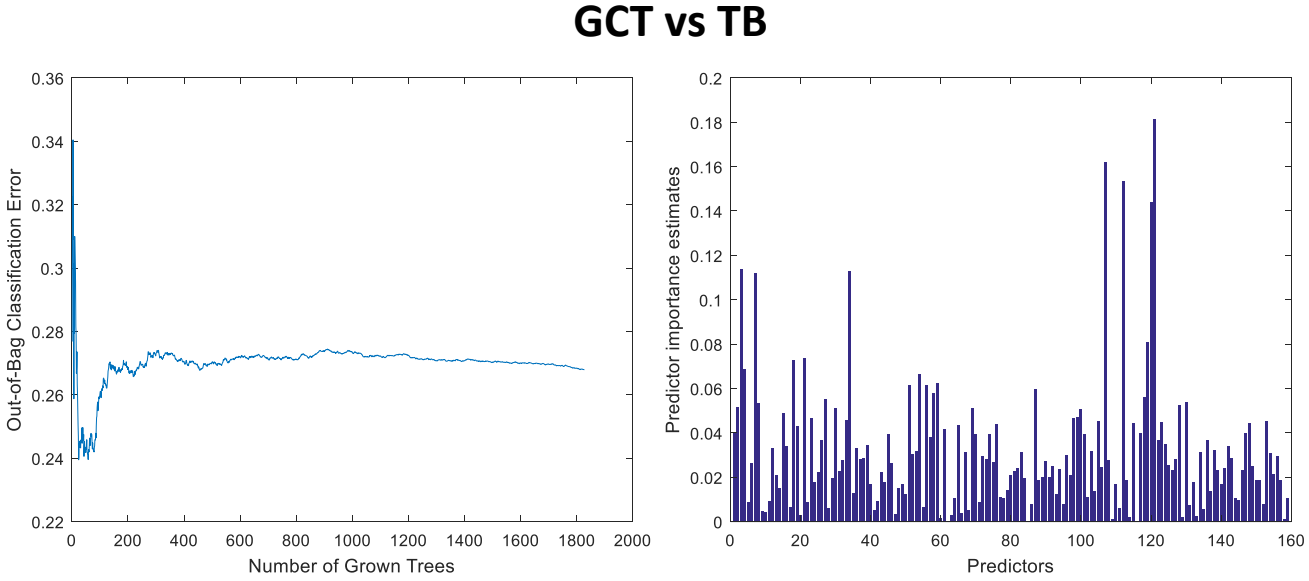


Figure 4.3.4 Classification Error and predictors' importance estimates between GCT and TB. Tree number was set as 1828.

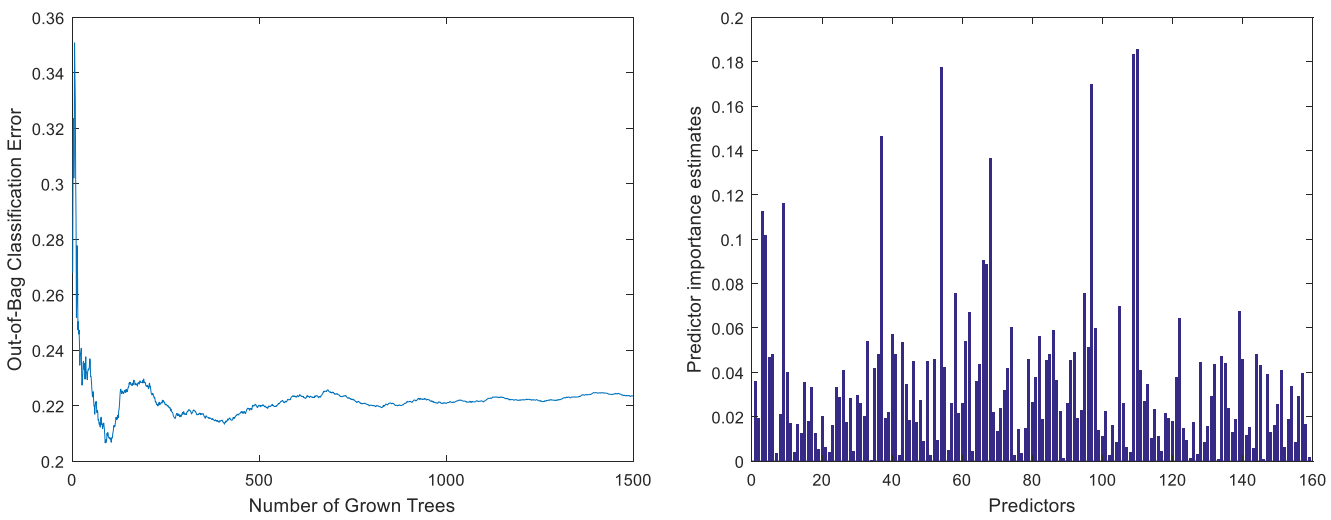


Figure 4.3.5 Classification Error and predictors' importance estimates between TB and Granuloma. Tree number was set as 1500.

After using random forest algorithm to find out 10 of most significance parameters, utilizing different combinations of these 10 features to train a logistic model to serve as a classifier is the next step. Different combination of features yield different model, so for sure the classification result will not be the same. After the model had been generated, ROC analysis with leave-one-out classification method was utilized to find the optimal diagnostic features. Due to the low number of case number, overfitting occurred during differentiation process and thus, number of features needed to be carefully selected. **Table 4.3.1** is the AUC result of 10 different combination among three types of spinal cancer. Group 1-3 is GCT vs Granuloma, GCT vs TB, TB vs Granuloma and red numbers are the best two results within the group. **Figure 4.3.5** is a more direct way to see the best two result in ROC curve.

Table 4.3.1 AUC result of 10 different combination among three types

	1:2	1:3	1:4	1:5	1:6	1:7	1:8	1:9	1:10
1	**0.7833**	**0.8111**	**0.7778**	**0.7833**	**0.8722**	**0.7306**	**0.8278**	**0.8111**	**0.7888**
2	**0.6556**	**0.7111**	**0.6333**	**0.7222**	**0.7167**	**0.6500**	**0.6389**	**0.5833**	**0.5611**
3	**0.6944**	**0.8958**	**0.8472**	**0.8264**	**0.7639**	**0.6319**	**0.4444**	**0.6250**	**0.5833**

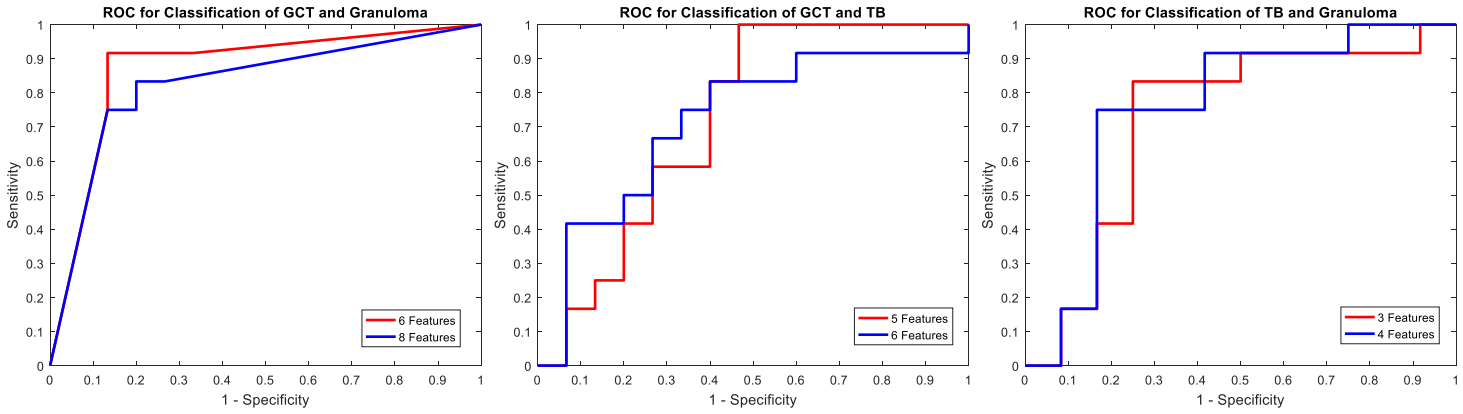


Figure 4.3.6 ROC curve of classification result among three types.

4.4 Summary and Discussion

For breast cancer machine learning, the morphology and texture features were firstly extracted from the segmented lesion mask. Random forest algorithm was then applied to choose out the most significance parameters that can differentiate well among three subtypes of breast cancer. Cross validation method was utilized when classification model were trained. Although the results of morphology and texture analysis for differentiating

among the three molecular subtypes were not satisfactory, this could be due to the small number of cases. In addition, three subtypes of breast cancer can differentiate well based on their biomarker, which may be difficult to show much difference in morphological and texture features.

In spinal cancer machine learning study, seven types of spinal cancer were included. Although metastatic cancer has the largest amount of case number and should contributed the most in this study, due to the high varieties of lesion characteristics, exclusion of this type is necessary to ensure the accuracy of the classifier. In the six left types, three cancers with highest case number were chosen to perform this study. Based on the segmented ROI from Chapter 3, totally 150 texture features were extracted from wash-out, wash-in and max enhancement maps for diagnostic classification. After validated by leave-one-out method, the AUC are around 0.72 – 0.90. The classification results can be improved if more cases are included in the analysis. More advanced classification method, such as graphic model and support vector machine, may be utilized in the future. Also, the proposed image segmentation methods may be further improved to enhance the diagnostic accuracy.

Chapter 5 Summary

With the continuously increasing prevalence of breast cancer and the importance to overcome difficulties in segmenting non-mass tumors in breast cancer MRI, due to the abnormal morphology of non-mass tumors, it's getting more and more clinically meaningful to propose well-behaved non-mass tumor segmentation methods. Also, it has been tough task for researchers to determine the locations of spinal tumors since both axial and sagittal MRI images have their defects, rendering it complicated in spinal lesion diagnosis.

In chapter 2, a new method for automatic non-mass lesions in breast MRI based on region growing was proposed, which was rarely reported in previous researches. This template-based approach turned out to be satisfactory since the overall correction rate of human validation is quite low. The quality of non-mass lesion segmentation and strength in time consumption makes this method promising in further use of MRI image processing. Since this is a relatively novel method for non-mass tumor segmentation in breast cancer, and it did had good behavior throughout our analysis based on amount of case numbers, we believe it could lead to the presence of more similar methods aiming at segmenting tumors in breast cancer and streamlining the analysis of breast cancer MRI images, and therefore making further contribution to the diagnosis and treatment of this disease.

We also combined sagittal-axial view registration, normalized cut and region growing in the spinal cancer lesion segmentations in chapter 3. Since spinal cancer is a major kind of cancers with strong variability, it's crucial to segment, locate and determine the shape of spinal tumors and differentiate spinal tumors in different types of spinal cancers. Seven types of spinal cancers were included in this study and most results turned out promising.

Although it was computational expensive for some cases in this study and unclear boundaries in MRI images might influence the final results of spinal lesion segmentation, this research gave us a direction in spinal tumor analysis and made it bright with regard to the challenges in future improvement.

Additionally, machine learning algorithm was used in this study. For breast cancer research, a classifier using 11 morphology features and 50 texture features was built to differentiate three subtypes of breast cancers. Possible reasons for unsatisfactory performance of this classifier might include the lack of abundant case number. Also, it would be better to differentiate these three subtypes of breast cancers with certain biomarkers, which might not be directly related to the morphology and texture features we extracted. Such problem apply to spinal cancer machine learning as well, where more cases and more advanced classification methods might improve the performance of the classifier. Although defects of the machine learning algorithm in this study are obvious, we did see possibilities of improvement. Algorithms based on other programming languages (Ex. R) might be helpful in boosting the speed as well bring classifier with better performance.

Reference

Chapter1 Introduction

1.1 "Magnetic Resonance, a critical peer-reviewed introduction". European Magnetic Resonance Forum. *Retrieved 17 November 2014*.

1.2 Humberstone, Miles R., and Guy V. Sawle. "Functional magnetic resonance imaging in clinical neurology." *European neurology* 36.3 (1996): 117-124.

1.3 Su, Min-Ying, and Jeon-Hor Chen. "Quantitative Measurements of Tracer Transport Parameters Using Dynamic Contrast-Enhanced MRI as Vascular Perfusion and Permeability Indices in Cancer Imaging." *Quantifying Morphology and Physiology of the Human Body Using MRI*. Taylor & Francis, 2013. 329-353.

1.4 Su, Min-Ying, et al. "Tumor characterization with dynamic contrast–enhanced MRI using mr contrast agents of various molecular weights." *Magnetic resonance in medicine* 39.2 (1998): 259-269.

1.5 Choyke, Peter L., Andrew J. Dwyer, and Michael V. Knopp. "Functional tumor imaging with dynamic contrast-enhanced magnetic resonance imaging." *Journal of Magnetic Resonance Imaging* 17.5 (2003): 509-520.

Chapter 2 Mass and non-mass tumor segmentation in breast cancer

2.1 Howell, Anthony, et al. "Risk determination and prevention of breast cancer." *Breast Cancer Research* 16.5 (2014): 446.

2.2 Goebl, Sebastian, et al. "CAD-system based on kinetic analysis for non-mass-enhancing lesions in DCE-MRI." *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2013.

2.3 Ertaş, Gökhan, et al. "Breast MR segmentation and lesion detection with cellular neural networks and 3D template matching." *Computers in biology and medicine* 38.1 (2008): 116-126.

2.4 Vag, Tibor, et al. "Kinetic analysis of lesions without mass effect on breast MRI using manual and computer-assisted methods." *European radiology* 21.5 (2011): 893-898.

2.5 Lin, Muqing, et al. "Template-based automatic breast segmentation on MRI by excluding the chest region." *Medical physics* 40.12 (2013).

2.6 Fwu, Peter T., et al. "Quantification of Regional Breast Density in Four Quadrants Using 3D MRI—A Pilot Study." *Translational oncology* 8.4 (2015): 250-257.

2.7 Chuang, Keh-Shih, et al. "Fuzzy c-means clustering with spatial information for image segmentation." *computerized medical imaging and graphics* 30.1 (2006): 9-15.

2.8 Albrecht, Andreas, and Gary Bernstein. "Evaluating dark energy probes using multidimensional dark energy parameters." *Physical Review D* 75.10 (2007): 103003.

2.9 Kirbas, Cemil, and Francis KH Quek. "Vessel extraction techniques and algorithms: A survey." *Bioinformatics and Bioengineering, 2003. Proceedings. Third IEEE Symposium on*. IEEE, 2003.

Chapter 3 Spinal cord cancer segmentation

3.1 Sundaresan, Narayan, et al. "Spinal chordomas." *Journal of neurosurgery* 50.3 (1979): 312-319.

3.2 Sanjay, B. K., et al. "Giant-cell tumours of the spine." *Bone & Joint Journal* 75.1 (1993): 148-154.

3.3 Eimoto, Tadaaki, et al. "Plasma cell granuloma (inflammatory pseudotumor) of the spinal cord meninges." *Cancer* 41 (1978): 1929-1936.

3.4 Li, M. H., Stig Holtås, and E-M. Larsson. "MR imaging of spinal lymphoma." *Acta Radiologica* 33.4 (1992): 338-342.

3.5 Black, Perry. "Spinal metastasis: current status and recommended guidelines for management." *Neurosurgery* 5.6 (1979): 726-746.
3.6 Garrett, Michael J. "Spinal myeloma and cord compression—Diagnosis and Management." *Clinical radiology* 21.1 (1970): 42-46.
3.7 Pertuiset, Edouard, et al. "Spinal Tuberculosis in Adults: A Study of 103 Cases in a Developed Country, 1980-1994." *Medicine* 78.5 (1999): 309-320.
3.8 Garg, Ravindra Kumar, and Dilip Singh Somvanshi. "Spinal tuberculosis: a review." *The journal of spinal cord medicine* 34.5 (2011): 440-454.
3.9 O'Connor, J. P. B., et al. "Dynamic contrast-enhanced imaging techniques: CT and MRI." *The British journal of radiology* (2014).
3.10 Shi, Jianbo, and Jitendra Malik. "Normalized cuts and image segmentation." *IEEE Transactions on pattern analysis and machine intelligence* 22.8 (2000): 888-905.
3.11 Carballido-Gamio, Julio, Serge J. Belongie, and Sharmila Majumdar. "Normalized cuts in 3-D for spinal MRI segmentation." *IEEE transactions on medical imaging* 23.1 (2004): 36-44.
3.12 Fan, Jianping, et al. "Seeded region growing: an extensive and comparative study." *Pattern recognition letters* 26.8 (2005): 1139-1156.
3.13 Adams, Rolf, and Leanne Bischof. "Seeded region growing." *IEEE Transactions on pattern analysis and machine intelligence* 16.6 (1994): 641-647.

Chapter 4 Machine Learning

4.1 Michie, Donald, David J. Spiegelhalter, and Charles C. Taylor. "Machine learning, neural and statistical classification." (1994).
4.2 Langley, Pat, and Herbert A. Simon. "Applications of machine learning and rule induction." *Communications of the ACM* 38.11 (1995): 54-64.
4.3 Niu, Lili, et al. "Surface roughness detection of arteries via texture analysis of ultrasound images for early diagnosis of atherosclerosis." *PloS one* 8.10 (2013): e76880.
4.4 Carlson, Joel. “Texture Analysis.” 04 May 2016. < ftp://cran.r-project.org/pub/R/web/packages/radiomics/vignettes/TextureAnalysis.html >
4.5 Liaw, Andy, and Matthew Wiener. "Classification and regression by randomForest." *R news* 2.3 (2002): 18-22.
4.6 Malhotra, Gautam K., et al. "Histological, molecular and functional subtypes of breast cancers." *Cancer biology & therapy* 10.10 (2010): 955-960.
4.7 Nie, Ke, et al. "Quantitative analysis of lesion morphology and texture features for diagnostic prediction in breast MRI." *Academic radiology* 15.12 (2008): 1513-1525.

Appendix

3D model movie of segmented non-mass tumor.