# Computer Aided Diagnosis System for Detection of Focal Cortical Dysplasia Lesions on T1 and T2 Weighted MRI

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Chin-Ann Yang

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

Advisor: Mostafa Kaveh

September, 2012

# Acknowledgements

I received tremendous help from a lot of people through the journey of earning my PhD. Among all of the them, I would like to first dedicate my greatest gratitude to my academic advisor, Prof. Mostafa Kaveh from University of Minnesota, Twin Cities. While being a very knowledgable signal processing expert and a great mentor, he also cares a lot about his students in both academically and personally. I not only learned a lot of technical skills from him, but also received a lot of useful suggestions from him to guide me through the ups and downs of studying aboard. I am very lucky to have him as my advisor.

I would also like to express my gratefulness to my co-advisor Dr.Bradley J. Erickson from the Mayo Clinic. The thesis was made possible largely due to the new established collaboration with the Mayo Clinic, Rochester. Many thanks to Dr. Erickson on being tremendously supportive during the course of the project. If not for his help on providing insightful insights and collecting clinical data, this project will not be possible.

Also, many thanks to the all the faculty members and staffs in the ECE department of University of Minnesota. Especially, I would like to thank Prof. Guillermo Sapiro, Prof. Emad Ebbini, and Prof. Bin He from the BME department, they are very kind to serve as my thesis committees. Thanks to their sharing of their expert knowledge and providing insightful input on the thesis, I am able to consider different approaches such as differential features and GrabCut. Here, I would also like to thank the ECE department which it regularly hold events for women graduate students to attend talks and activities. I especially appreciated the chance

to have discussions with well-established female scholars in the EE field, these amazing women inspire me in a lot of way. Thanks to the ECE department, I have the chance to talk and share ideas with them. I would also like to thank all of my female friends in the department, Jing Wang and Hao Zhu which I have developed kinder spirit through the past few years. Through the course of my research, there are also many others from the University of Minnesota who have helped me and shared valuable insights with me. I would like to thank my office-mate Jimeng Zheng, Wentao Shih, Bryan Perfetti and Dr. Hossein Zare for not only helping me academically but also being extremely supportive as friends. I would also like to thank Dr. Iman Aganj for generously providing his code of his excellent work on thickness measurement algorithm to assist the project.

Also, the Taiwanese students communities have play a key role in my US life. I would like to dedicate my appreciation to the Taiwanese Student Associations, for holding events such as Chinese New Year Party. Those events help me get together with a lot of fellow Taiwanese students and they have been my family here in Minnesota. Among all of my dearest friends, I would especially like to thank Chung-I Ho, who has been my greatest support for many years. He not only guided me through the most difficult time in my life, but also took me in as a family. I would always be thankful of him.

Most importantly, I would like to thank my whole family for their continuous support during my graduate study. My father, Jar-Ferr Yang encouraged me to become a Engineer since I was very young. He always has the faith in me and trust me that I can accomplish a lot of great things. My mother, Pi-Hsia Hung can always give me peace and joy whenever I talk to her, she has provided me with her warmest heart and that is the greatest motivator in the world for me to face any challenges. I understand my parents always have great expectations in me and I hope I can continue to make them proud in the future of my life. And also, I would like to thank Yi-shu Tai, who for now, is like part of my family. He has been nothing but supportive. His companionship is the power source that I need to continue my American dream. I am very lucky to have him in my life.

## Abstract

Focal cortical dysplasia (FCD) is the most frequent malformation for patients with pharmacoresistant epilepsy that require surgical treatment. Providing automated procedures to detect FCD lesions is greatly desirable because visual diagnosis is often challenging, time consuming, and relies highly on the individual's expertise. In this thesis, we propose two Computer Aided Diagnosis (CAD) approaches for Focal Cortical Dysplasia (FCD) lesion detection and segmentation on T1 and T2 weighted MRI. For the first CAD system, an automatic detection algorithm for FCD lesions on T1 weighted MRI is proposed. Instead of using the traditional voxel-based analysis, we introduce a set of volume-based statistical features with Naive Bayes Classifier. Subsequently, a set of cluster-based differential features with a Support Vector Machine (SVM) classifier is used to eliminate the false positives (FPs) resulting from the first processing stage. The advantage of our system lies on the use of volume-based analysis to allow the study of feature distributions in a spatial neighborhood. The second CAD system automatically segments FCD lesions on T2 weighted MRI. We present a Markov Random Field (MRF) model for the segmentation task with a particular emphasis on the incorporation of T1 information with a location prior. By integrating such location prior, we take the advantage of T1 weighted MRI in producing better differentiation of soft tissues into the T2 lesion segmentation task.

The proposed algorithms are validated on a dataset that consists a total of 51 subjects with FCD lesions provided by the Radiology Department of Mayo Clinic. The experimental results show a 87% FCD lesion detection rate for T1-weighted MRI and a 100% FCD lesion detection rate for T2 weight MRI. The experimental results also show that proposed methods outperform previous methods in the literature .

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

The goal of Computer Aided Diagnosis (CAD) is to facilitate clinical decisions based on quantitative analysis. With the recent advancements of computer technology, CAD has been an active research field that aims at resolve challenging diagnosis issues with engineering solutions. The engineering solutions usually include techniques from multiple disciplines such as image/signal processing, pattern recognition, machine learning, and statistical analysis. In this thesis, a CAD system for Focal Cortical Dysplasia (FCD) based on both T1 and T2 Magnetic Resonance Imaging(MRI) is proposed to assist radiologists in evaluating and diagnosing disease.

Originally, image-based analysis for CAD application has been only limited to using basic image processor elements(eg. intensity) or operators(eg. gradient) as features. While these can provide some partial information of the image contexts, it does not provide discriminative description of the image when it comes it medical image understanding for CAD applications. Recently, with the connection liked to pattern recognition, machine learning and statistics, several methods have been proposed to incorporate medical prior knowledge into the selection of features (eg.morphological-based features) and machine learning, which often lead to more successful results than general features and learning methods.

We start the introduction of the problem of FCD lesion detection and analysis

by presenting some example images and their imaging. This chapter was summarized with a example result from the work of thesis and a brief description of the contributions of the thesis, more details on the methods will be discussed in Chapter 3 and Chapter 4. In Chapter 2, a review of related work is given. In Chapter 5 and Chapter 6, we will describe the dataset used, experimental setup and present the experimental results. In Chapter 7, discussions, conclusions, and possible direction of future work are provided to conclude this thesis.

## 1.1 Data Example

One of the most popular approach for CAD algorithm is based on learning and generalizing from the data collections of the disease of interest. In this thesis, all the MRI images presented and analyzed are selected from the Radiology Department, Mayo Clinic, Rochester. Fig 1.1 shows an example of a T1-weighted 3D MRI scan of a subject with FCD. Fig 1.2 shows another example of a MRI scan of T2-weighted Fluid Attenuated Inversion Recovery(FLAIR) sequences. It is common practice to perform multiple modalities scan for a single subject, in order to allow radiologists or physicians to gather information taking advantages of different imaging modalities.

**The challenge**

FCD lesion detection is a difficult task even for trained experts. Research has shown the in a retrospective series of pathologically proven FCDs the MRI was reported as normal in 34% [1]. The major challenges for the FCD lesion detection problem are presented as follows:

1. *FCD lesions are subtle* FCD lesions often located at the boundary of the GM and WM junction and their intensities often resemble the intensities of the healthy GM. Therefore, FCD lesions are challenging even for trained experts, and more so to a computer program.

2. *FCD lesions have variations* FCD lesions vary greatly in intensity, shape, and texture. Although there are three characteristics as mentioned in [2], these characteristics may or may not coexist [3].

3. *Image quality, image inhomogeneity* The FCD dataset consists of image scans with variable image quality (eg.signal-to-noise ratio, patient motion artifacts) and resolutions (Magnetic field 1.5T and 3T). To design a successful CAD system for FCD detection, the computer program needs to be able to account for these varieties.

The approximation of human observers' perceptions have always been the holy grail of any pattern recognition problem. To design a CAD method, one should never aim to replace human's opinion, but to use computational power to assist radiologists or physicians in arriving at the best clinical decision, which is the goal of this thesis.

## 1.2 The Approach

The thesis can be divided into two main parts which consist of two CAD systems to perform FCD lesion detections on two different image modalities: T1 weighted MRI and T2 weighted FLAIR MRI. The two systems use different image processing and machine learning techniques tailored for the different FCD lesion characteristics on each specific modality. An example can be seen in Fig 1.3. The FCD lesion presented for the same subject with different imaging modalities has different image characteristics: intensity, texture, shape, etc. This thesis solves the two problems separately and then compares the results.

### 1.2.1 Computer Aided Diagnosis of FCD on T1 weighted MRI

FCD lesions on T1 weighted MRI have three major characteristics: increased cortical thickness, blurriness in the gray matter (GM) and white matter (WM)

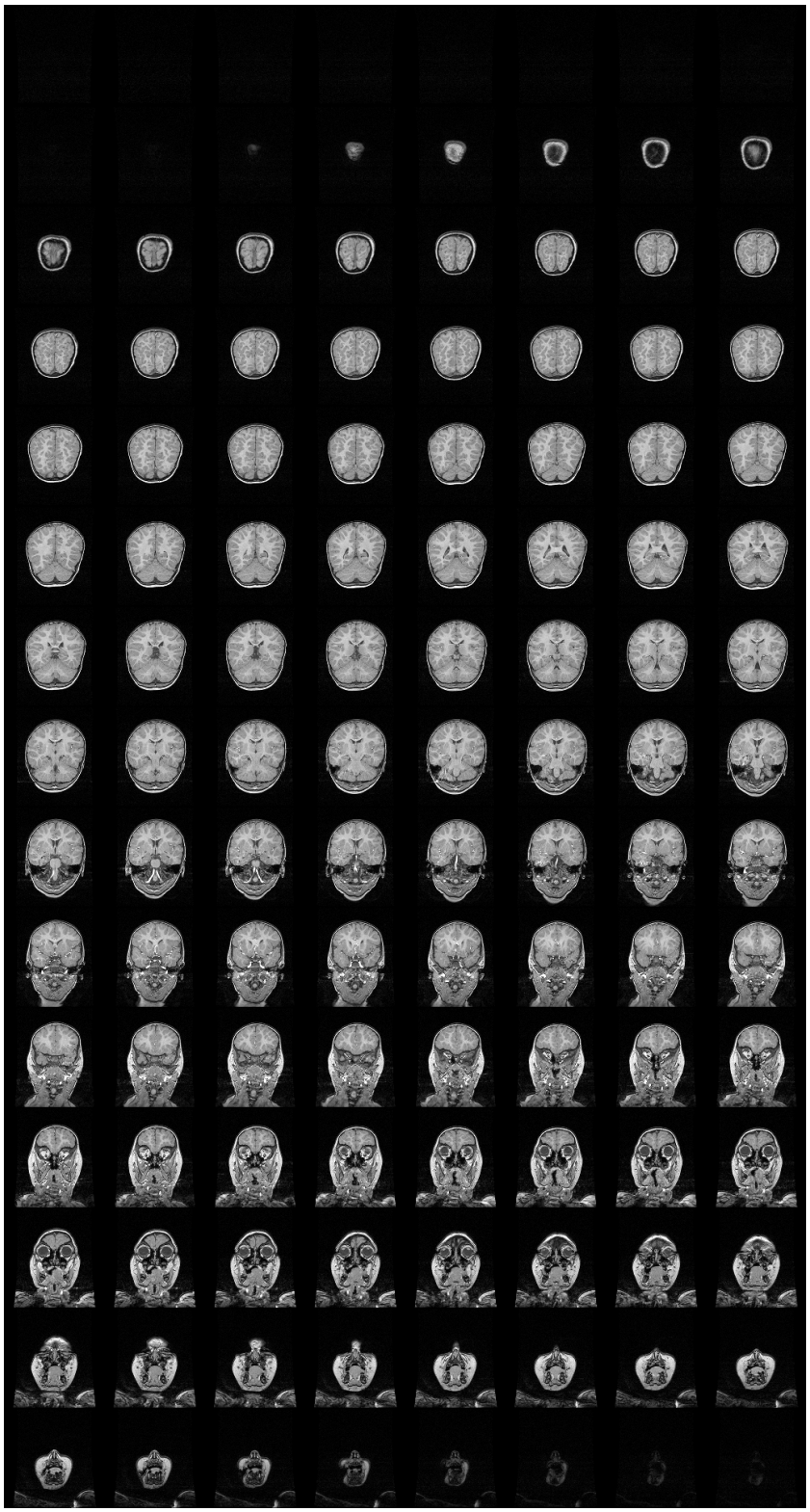

Figure 1.1: 3D T1-weighted MRI images with FCD lesions

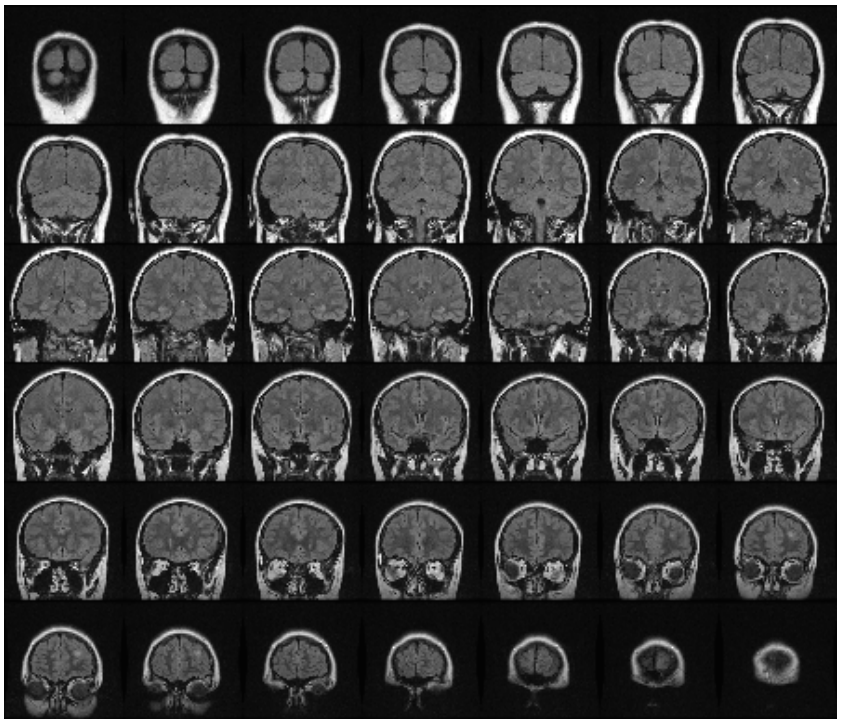

Figure 1.2: 3D T2-weighted MRI images with FCD lesions

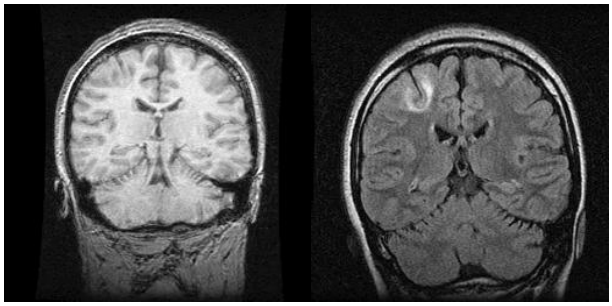

Figure 1.3: Focal Cortical Dysplasia on T1(left) and T2(right) -weighted MRI

junction, and changed signal in the cortex of the underlying WM. Our CAD system first preprocesses the image, and then divides the processed MRI volume into several volumetric cubes. For each cube, the proposed features are calculated. Based on these features, a Naive Bayes Classifier [4] is applied to classify each cube. Each cube is given a label according to the classification, either marked as lesional or healthy. Afterwards, the positive or lesional output from Naive Bayes classifier contains several clusters in the MRI volume domain. Another stage of analysis and classification is applied to the clusters by introducing cluster-wise features and using a Support Vector Machine (SVM) [5] classifier, which further classified each cluster as lesional or healthy. Last, the final flagging is determined and given to the radiologist for evaluation. The proposed CAD system for FCD lesion detection in a T1 weighted MRI image is summarized in Fig 1.4.

### 1.2.2 Computer Aided Diagnosis of FCD on T2 weighted FLAIR MRI

The FCD lesion on T2 weighted MRI has different characteristics than T1 MRI, in that a FCD lesion is typically demonstrated only by increased signal intensity [6]. With the intensity being the only feature, there will be false positives due to noise and intensity inhomogeneity. We incorporate two techniques to provide a more robust segmentation algorithm: (1) Markov Random Field modeling, (2) combination T1 location prior information. The CAD system for FCD detection and segmentation on T2 weighted MRI is summarized in Fig 1.5.

Here we provide a preview of the FCD lesion detection results. An example result is provided for each CAD system proposed. For the FCD detection on T1-weighted MRI, the lesion detection is presented in Fig 1.6.

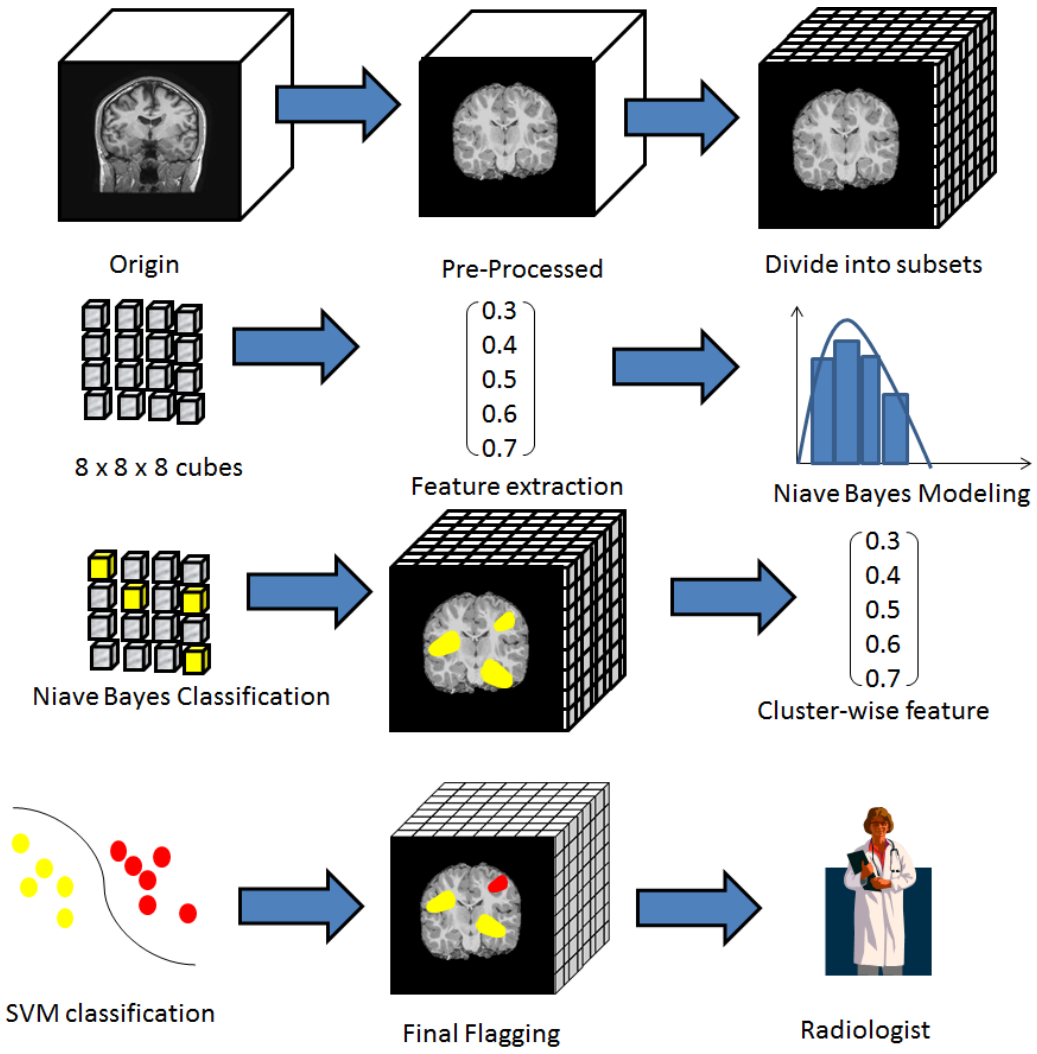


Figure 1.4: CAD system flow chart for FCD detection on T1 weighted MRI

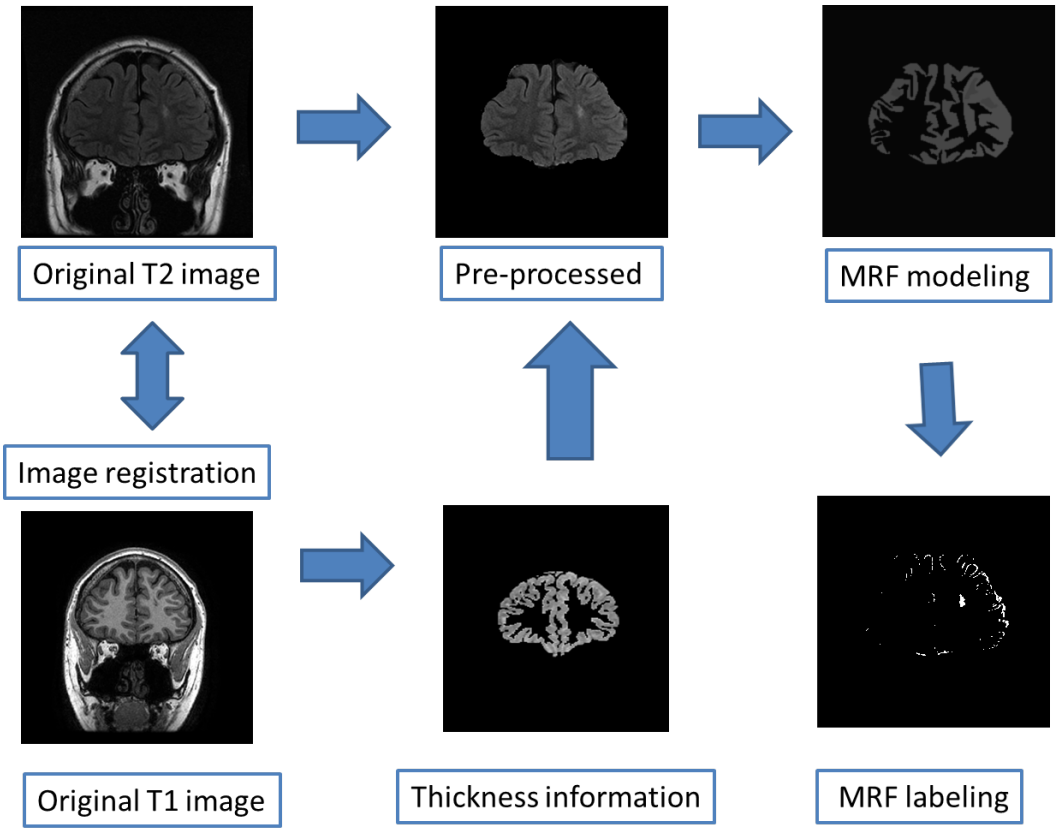


Figure 1.5: CAD system flow chart for FCD detection on T2 weighted MRI

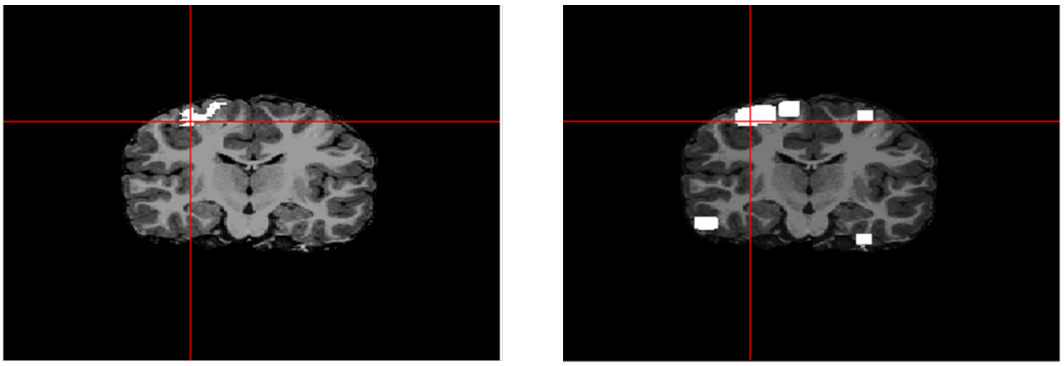

Figure 1.6: Example 1: FCD detection on T1 weighted MRI

## 1.3 Contributions

In this work, we present two CAD systems for image-based FCD lesion detections on T1 and T2 weighted MRI and validate these on 51 subjects with FCD lesions. In additions, to the demonstration of fully automatic frameworks to assist FCD lesion detection and diagnosis, the other major contributions include

1. development of new 3D volumetric cube-based features and cluster-wise features on FCD lesion detections on T1-weighted MRI. These methods provide a novel way to approach the problem;

2. development of the first fully automatic CAD system for FCD lesion detection on T2 FLAIR weighted MRI. The segmentation is achieved by incorporating location prior information to assist the segmentation.

3. robust validation by conducting experiments on a extensive FCD clinical data set with 51 subjects and the results are compared with the widely referred Antel's method [7];

4. comparison of automated results on T1 and T2, which provides a subjective indicator of the sensitivity of FCD automatic lesions detection task on each modality.

## 1.4 Outline

In the next chapter, we provide some background knowledge and present a review of existing methods of FCD lesion segmentation on both T1 and T2 images. The purpose of this chapter is to provide an overview of the commonly used methods in the literatures, as well as guidelines for choosing reasonable features and classifiers for our particular application. Chapter 3 explains our proposed features, methodologies and the classification algorithms to perform the FCD lesion detection task on T1 weighted MRI. Chapter 4 provides similar content as

in Chapter 3 for the CAD system proposed for FCD lesion segmentation with T2 FLAIR weighted MRI.

In Chapter 5, we provide various statistical analyses of the proposed features, and the performance of the classification results on the dataset which consists of 51 FCD subjects is presented in Chapter 6. In Chapter 7, the dissertation concludes with a discussion of the experimental results and future research directions which can be followed up with the results of this work.

# Chapter 2

# Background

In this chapter, we provide background knowledge that is related to the thesis. In Section 2.1 characteristics in different modalities of Magnetic Resonance Imaging(MRI) is briefly introduced. In Section 2.2, we discuss the formulation and significance of the type of epilepsy related to Focal Cortical Dysplasia (FCD). In Section 2.3, the basic of designing a Computer Aided Diagnosis (CAD) system is provided and Section 2.4 gives an overview and discussion of two general approaches of machine learning methods used in a CAD system. Finally in 2.5, we review the previous work on the FCD lesion detection and discuss their pros and cons.

## 2.1 Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is a medical imaging technique based on known physical phenomena called Nuclear Magnetic Resonance (NMR)[8] and Hydrogen nuclei are the basis for the signal. The resulting MR image is an image of the transverse magnetization of hydrogen nuclei at the time when the signal is detected. MRI has gained a widespread acceptance for a large variety of medical explorations. It is one of the most common techniques used in radiology to visualize the body's soft tissues, including the brain, because of its ability of providing

excellent contrast. There are several different types of MRI, such as functional MRI (fMRI) and diffusion MRI. However, in this PhD thesis, the discussions only include structural MRIs. Among structural MRIs, there are two types of basic MRI scans: T1-weighted and T2-weighted MRI. T1 and T2 weighted MRI provide complementary information and are beneficial to quantitatively analyze both modalities while making clinical decisions. For example, T1 weighted MRI provides excellent gray-matter (GM) and white-matter (WM) contrasts whereas cerebral-spinal-fluid(CSF) is presented best on T2-weighted MRI. The basics and differences of T1 and T2-weighted MRIs are summarized as follows [9]:

*T1-weighted MRI* T1 weighted images can be obtained by setting short repetition time (TR) ($< 750ms$) and echo time (TE) ($< 40ms$) values in conventional spin echo sequences, while in Gradient Echo Sequences can be obtained by using flip angles of larger than 50° while setting TE values to less than 15 ms. The intensity values of T1 are significantly different between grey matter and white matter and is widely used when undertaking brain scans.

*T2-weighted MRI* T2 weighted images can be obtained by setting long repetition time (TR) ($> 1500ms$) and echo time (TE) ($> 75ms$) values in conventional spin echo sequences, while in gradient echo sequences they can be obtained by using flip angles of less than 40° instead of 90° while setting TE values to above 30 ms. This also makes it more prone to losses at air/tissue boundaries, but can increase contrast for certain types of tissue, such as venous blood.[9]

Also, there are some specialized MRI scans. Here we discussed the Fluid Attenuated Inversion Recovery (FLAIR) pulse sequences. FLAIR sequences have been developed to null the signals from cerebrospinal fluid (CSF) and yield heavily T2-weighted images. FLAIR has been especially useful in diagnosing central nervous system diseases [10]. By carefully choosing the inversion time TI (the time between the inversion and excitation pulses), the signal from any particular tissue can be suppressed.

In this section, the background of the two modalities of MRI: T1 and T2 FLAIR MRI was introduced. In this thesis, we present two CAD systems to solve the Focal Cortical Dysplasia (FCD) lesion detection tasks on both modalities. In the next section, we provide a brief introduction on FCD and discuss its clinical importance.

## 2.2 Focal Cortical Dysplasia

Epilepsy is one of the most common chronic neurologic disorders that is characterized by seizures. Epilepsy affects fifty million people in the world and it is estimated that up to 22.5% of patients with epilepsy have medically refractory (drug resistant) epilepsy [11]. With drug-resistant epilepsy, only surgical removal of the associated lesions can help improve the patients' quality of life, and studies have shown that surgery results in significant reduction of seizures if the entire lesion is resected [12][13]. With the recent advances in imaging technology, lesions can be more accurately identified and thus more amenable to resection. Among all of the imaging modalities, high resolution Magnetic Resonance Imaging (MRI) is the most widely used for the diagnosis of epilepsy, having sensitivity of 63% to 98% [14]. The major advantages of MRI is the topographical characterization of the lesion with respect to its size, location and extension [15].

Focal Cortical Dysplasia (FCD) is the most common cause of medically refractory epilepsy. However, FCD diagnosis can be challenging and time-consuming even for expert radiologists due to only subtle structural changes can be detected on the images, even with high resolution MR imaging. Studies have indicated that the ability to define and fully excise the entire region of dysplastic cortex is the most powerful variable influencing outcome in patients with FCD [13]. Hence, there is a demand for automatic and objective procedures to assist radiologists in FCD detection and quantification. In 1971, Taylor et al[2] characterized Taylor-type FCD which is also known as Type 2 FCD. Taylor-type FCD is determined histologically due to certain features such as giant neurones, dysmorphic neurones

with or without balloon cells. In this thesis, we restrict our discussion on Taylor-type FCD, which has three main characteristics on T1-weighted MRI: increased cortical thickness [16][17], blurriness in the grey matter (GM) and white matter (WM) junction [18] [17][19], and changed signal in the cortex and underlying WM [20][21]. On T2-weighted MRI, FCD is demonstrated by increased signal intensity [6].

Detecting and analyzing dysplastic lesions on MRI images plays an important role in providing treatment planning and serves as a road map to surgeons for presurgical evaluations. Currently, the diagnosis and analysis are still performed manually, the radiologist usually takes up 30 minutes per study and only indicates where lesions locate, but does not quantify the size or extent [22] of the lesion. Although the current procedure is used as a standard clinical routine, there are two major limitations: (a)it is not quantitative (b)three dimensional analysis is limited. A computer aided diagnosis (CAD) system has the capability to perform the task without these limitations, and can therefore assist radiologists in FCD detection and quantification. In the next section, we briefly introduce the basic methods and formulations in the existing computer aided diagnosis (CAD) systems and their role in the clinical practice.

## 2.3 Computer Aided Detection and Diagnosis

CAD or quantitative medical image analysis is a broad concept and field of research that integrates image processing, computer vision, machine learning and statistics into computerized procedures that assist radiologists in their clinical decision-making process. Such techniques include the detection of disease or abnormalities, the classification of lesions, the quantification of disease, risk assessment, and physiological evaluation. The methods can be applied to a wide-range of imaging modalities such as X-ray, computed tomography (CT), MRI, and ultrasound; with a target region of different part of the body (brain, breast [23], cardiac, etc.). These techniques seek to maximize the information that may be

extracted from medical images to augment radiologists' subjective, qualitative interpretation of the displayed images with objective, quantitative results obtained from numerical analysis of the image data. and have received much attention for the past two decades largely due to the improved fidelity and resolution of medical imaging systems[24]. A typical CAD, which is fundamentally equivalent to a highly complex pattern recognition task, consists of these following methodologies:

- **Preprocessing** This step is generally necessary for removing noise and condition the image or signal qualities by noise filtering, leveling intensities or changing contrast parameters.

- **Segmentation** This step differentiates different structures in the images for further analysis. (e.g. heart, lung, ribcage) For brain images, this step can help remove the skull or cerebellum that are unnecessary for the further evaluation in the CAD system. This step can be achieved by using segmentation method such as Template Matching.

- **Structure/ROI Analysis** This step analyzes regions individually for special characteristics. The characteristics could be the form, size, location, or intensity. Some human-extracted or computer-extracted features can also be used for analysis. Computer-extracted features are objective and reproducible ,therefore are more desirable.

- **Evaluation/Classification** After the structure is analyzed and the features are selected, every ROI is evaluated individually for the probability of True Positive(TP). Some of the procedures include:

  - Nearest-Neighbor Rule
  - Minimum distance classifier
  - Cascade classifier
  - Bayesian classifier
  - Radial Basis Function(RBF)

– Support Vector Machine (SVM)

The final decision is based on the output of the classification method. For example, for the Bayesian classifier, if the detected structures have reached a certain threshold level (e.g. the probability of being TP is larger than 0.8), they are highlighted in the image for the radiologist for further examination. CADs help attract the examiner's attention to subtle features that might have been overlooked or dismissed as normal, and also assist the radiologists to process large amount of information more efficiently. CADs has a lot of applications and have demonstrate great potential impact on clinical trials[25].
In this thesis, we develop two new CAD algorithms that are automatic and reproducible. The work is designed specifically for detecting Focal Cortical Dysplasia (FCD) lesions on T1 and T2 brain MRI.

## 2.4 Overview of Machine Learning methods in CAD application

As we mentioned in the previous section, the final stage of designing a CAD system is based upon a selection of classification or machine learning methods for decision making. The selection of such method can play a major role in determining the quality of the CAD system. In this section, we give an overview of some of the most widely used machine learning methods for CAD application and discuss the pros and cons for these techniques.

For any given data set or input $x$, a machine learning method seeks to obtain a label $y$ that is optimal based on previously learnt examples. In other words, machine learning methods want to learn $p(y|x)$ and and choose the most probable label $y$. From Baye's rule, we know that

$$p(x, y) = p(y|x)p(x) = p(x|y)p(y). \tag{2.1}$$

Machine learning methods can be divide into two main categories based on

how they arrive at the solution of $p(y|x)$, which are Generative and Discriminative methods. Generative methods involves modeling of the training data set, on the other hand, discriminative methods performs the classification directly. In the following, we briefly introduce the general concepts, advantages, disadvantages and examples for both methods.

### 2.4.1 Generative Methods

For a given data set with input $x$ and its label $y$, generative classifiers learn a model by modeling the joint probability $p(x, y)$ and prior probability $p(y)$ and likelihood function $p(x|y)$. These are subsequently used to evaluate the posterior probability $p(y|x)$. Generative probabilistic models provide a rigorous platform to define the prior knowledge of the problem that is available from human experts, and combine it with the observed data. Based upon the knowledge based setting, generative methods have great success in the medical diagnosis field. Their ability to model uncertainty while applying prior knowledge is one of its advantages [26]. Some popular models for generative methods include: Gaussian, Naive Bayes, Mixtures of Multinomials, Mixtures of Gaussians, Mixtures of experts, Hidden Markov Models (HMM), Sigmoidal belief networks, Bayesian networks and Markov random fields. In general, generalization performance of generative models is often found to be worse than those of discriminative models due to differences between the model and the actual distribution of the data [27][28]. The advantages and disadvantages of generative methods are summarized as follows:

**Advantages of Generative Methods**

1. *Ability to introduce prior knowledge*
   The ability for generative model to introduce prior knowledge lies in the fact that its working in the joint distribution space, which allows inserting prior knowledge about relationships between variables and prior distributions.

2. *Do not require large training sets*

If the training sets are a good representation of the overall data set, and the actual distribution the data is in agreement with the model assumption, generative methods often require only small quantities of labeled data to classify large quantities of unlabeled data. This can be valuable when the labeling of data is not available or expensive.

**Disadvantages of Generative Methods**

1. *Marred by generation optimization criteria*
   The performance of generative methods often depend on parameter estimation of a given model, which is marred by the generation optimization criteria.

2. *Potentially wasteful modeling*
   Generative methods model the data without consideration of the classification task, this could lead to wastefully model variability which is unimportant for classification

3. *Reliant on domain expertise*
   Domain knowledge is often required to select appropriate models and parameters for generative methods.

4. *Don not scale well to large number of classes*
   Generative methods are traditionally only applied to smaller number of classes for its lack of discrimination ability.

**Generative method example: Naive Bayes Classifier**

A Naive Bayes classifier is a probabilistic classifier that utilizes Baye's Theorem with the assumption that all of its features are statistically independent. Naive Bayes approach is computationally efficient, and despite the fact that the naiveness assumptions are often not true, this classifier has worked well in many complex real-world situations [4]. Assume the attributes $X_1, X_2, ..., X_n$ are all conditionally

independent of one another given the class or outcome $Y$. Then the representation of $P(X|Y)$ can greatly be simplified as,

$$P(X_1, X_2, ..., X_n|Y) = \prod_{i=1}^{n} P(X_i|Y). \tag{2.2}$$

We can use this to derive the prediction, or the posterior probability $P(Y|X)$. The expression for the probability that $Y$ will be taken on its $k$th probable value, according to Bayes rule, is

$$P(Y = y_k|X_1, ..., X_n) = \frac{P(Y = y_k)P(X_1, ..., X_n|Y = y_k)}{\sum_j P(Y = y_j)P(X_1, .., X_n|Y = y_j)}, \tag{2.3}$$

where the sum is over all possible values. By assuming conditional independence,

$$P(Y = y_k|X_1, ..., X_n) = \frac{P(Y = y_k)\prod_i P(X_i|Y = y_k)}{\sum_j P(Y = y_j)\prod_i P(X_i|Y = y_j)}. \tag{2.4}$$

Since the denominator does not depend on $k$, we can have the Naive Bayes prediction rule based on the maximum a posterior (MAP) probability simplifies to:

$$\arg\max_{y_k} P(Y = y_k)\prod_i P(X_i|Y = y_k). \tag{2.5}$$

In this section, we discuss the generative approach and its pros and cons. An example method, Naive Bayes is introduced. Generative approach has its advantages of combing prior knowledge, however, when it comes to discrimination power, the discriminative approach has superior performance. The discriminative approach is discussed in the next section.

### 2.4.2 Discriminative Methods

Discriminative classifiers model the posterior $p(y|x)$ directly. As opposed to generative methods, discriminative methods do not focus on finding a solution using prior knowledge. On the contrary, it is almost impossible to embed prior knowledge into discriminative methods, which makes discriminative methods often being referred as a "block box": A data-point is given as an input, and $p(y|x)$ is returned, but without a clear understanding of how or why [26].

Instead of trying to figure out which distribution the data came from, the discriminative model only concentrates on approximating the shape of the boundary between classes. Some popular models include: Logistic regression [29], Support Vector Machines (SVMs), traditional neural networks, k-Nearest Neighbor (kNN) and Conditional Random Fields (CRF)[30] and their advantages and disadvantages are described as follows.

**Advantages of Discriminative Methods**

1. *Fast predication speed*
   Discriminative methods can make fast predictions for new (test) data points once the classifiers are trained, while generative models often require iterative solution.

2. *Potentially more accurate prediction*
   discriminative methods are generally expected to have better predictive performance since they are trained to predict the class label. On the other hand, generative methods are only trained to model the distributions of the input data.

**Disadvantages of Discriminative Methods**

1. *Task specific*
   Discriminative methods are very task specific, therefore, it could be difficult to apply previous trained discriminative model to a similar task.

2. *Long training time*
   Discriminative methods are known for their prediction performance, however, the more sophisticated algorithms often associate more training time.

3. *Do not easily handle compositionality* [31]
   Generative models can readily handle compositionality(e.g. faces with glasses

and/or hats, and/or moustaches)whereas standard discriminative models need to see all combinations of possibilities during training.

**Discriminative method example: Support Vector Machines**

In this section, we introduce a classic example for discriminative methods called the Support Vector Machines (SVMs). SVMs [5] are popular machine learning tools that have demonstrated successful results on practical pattern recognition problems. It formulates the machine learning problem as an optimization problem. In a binary classification setting such as our case in this study (lesional, healthy), the aim is to find a separating hyperplane between the two datasets that maximizes margins between the two classes. With training examples of the form: $\{x_i, y_i\}$, $i = 1, ..., n$ and $x_i \in \mathbb{R}^d$, $y_i \in \{-1, +1\}$ to denote class label. We can use parameters $w$,$b$ and write our classifier as,

$$h_{w,b}(x) = g(w^T x + b). \tag{2.6}$$

Here, $g(z) = 1$ if $z >= 0$ and $g(z) = -1$ otherwise. One can observe that for our prediction to be correct, we would like the functional margin $\hat{\gamma}^{(i)} = y^{(i)}(w^T x+b) > 0$ and hope the functional margin to be as large as possible to represent our confidence in the correct prediction. We denote $\hat{\gamma}$ as the smallest of the functional margins among all the training examples. With some manipulations, the classification problem can be formulated into the following optimization problem:

$$\min_{\gamma,w,b} \frac{1}{2}\|w\|^2 \tag{2.7}$$
$$\text{s.t. } y^{(i)}(w^T x^{(i)} + b) \geq 1, i = 1, ..., m.$$

This optimization problem is a Quadratic Programming (QP) problem and can, therefore, be efficiently solved. This is the primal problem for SVM. By utilizing Lagrangian Multiplier and KarushKuhnTucker (KKT) conditions, the dual optimization problem is,

$$\max_{\alpha} W(\alpha_1) - \frac{1}{2}\sum_{i,j=1}^{m} y^{(i)} y^{(j)} \alpha_i \alpha_j (x^{(i)})^T x^{(j)} \tag{2.8}$$

$$\text{s.t. } \alpha_i \geq 0$$

$$\sum_{i=1}^{m} \alpha_i y^{(i)} = 0.$$

The above formulation is the original SVM and a linear classifier. However, when the training instances are not linearly separable, one can perform the *kernel trick*. The resulting algorithm is formed similarly, except that every dot product is replaced by a nonlinear kernel function. This allows the algorithm to fit the maximum-margin hyperplane in a transformed feature space. The transformation may be nonlinear and the transformed space is high dimensional; thus, though the classifier is a hyperplane in the high-dimensional feature space, it may be nonlinear in the original input space. Some common kernels include:

- Polynomial(homogenous): $k(x_i, x_j) = (x_i \cdot x_j)^d$
- Polynomial(inhomogenous): $k(x_i, x_j) = (x_i \cdot x_j + 1)^d$
- Gaussian radial basis function: $k_(x_i, x_j) = exp(-\beta \|x_i - x_j\|^2)$
- Sigmod functions: $k(x_i, x_j) = tanh(kx_1 \cdot x_j + c)$

In this thesis, we utilize a SVM with polynomial kernel ($d = 3$) in Section 3.5 to classify cluster-based features that are described in Chapter 3. In the next section, we review the previous proposed CAD systems. For all of the CAD systems in the literature, it is impossible for one to complete an overview on all systems. Therefore, in the following section, we especially direct our attention on systems designed for the detection of FCD lesions or lesions of similar neural disorders.

## 2.5 Related Work

### 2.5.1 FCD lesion detection on T1-weighted MRI

For the demand of quantitative and subjective evaluation and detection of FCD lesions [32], several research efforts have been dedicated to the detection [7][33][34][35][19],

enhanced visualization [36][37][38][39][33][40], segmentation or delineation [41][42][43], and evaluation or analysis [43][37][41][44] of FCD lesions on T1 weighted MRI using quantitative image processing operators. In 1999, one approach proposed by Bastos et al. used a curvilinear multiplanar reformatting of the three dimensional MRI data to help improve visual diagnosis[39]. A more recent work [33] also proposed an automatic curvilinear reformatting method, which does not require user intervention based on a more sophisticated graph-based approach to segmentation. These approaches have been useful in assisting diagnosis by providing different perspectives for the radiologists. However, these still require knowledge and large amount of time of a trained expert to determine the location of the lesion, and also, their performances are highly based on the preprocessing segmentation results, which can be difficult with the presence of lesions. In this study, we will focus our discussion mainly on automatic and quantitative methods the requires minimum human intervention. Several reports [36][33][37][7][45][42][46][34] have appeared on the subject of automated detection procedures of FCD lesions. The proposed methods can be divided into three categories based on the main methodologies they use for the detection of lesions or abnormalities: *asymmetry-based analysis* [42][46], *template-based analysis* [19][34], and *discriminative feature-based analysis*[36] [37] [7] [45][42].

*Asymmetry-based analysis* methods utilize the fact that the human brain is highly symmetrical with respect to the mid-sagittal plane (MSP), and focus on detecting abnormalities based on the differences between the symmetrical locations. Asymmetry-based methods provide an inner-case, registration-free analysis. However, the main challenge here is that human brains are almost never perfectly symmetric [47]. Also, for subtle lesional structures like FCD, the differences in symmetrical locations may not be prominent enough for automatic detection.

*Group comparison methods* can also be called template-based methods which aim at detecting abnormalities by assessing differences between subjects' volume

data and a template constructed from healthy controls. These methods are more able to preserve spatial and structural information compared to other type of methods that usually divide the brain into smaller subsets, yet they are sensitive to the results of spatial normalization and registration, and may not be suitable for children subjects who have much smaller brain sizes compared to the template. Several group comparison methods have been very popular within the medical image processing communities, which include the most popular approach: Voxel-based morphometry (VBM) with Statistical Parametric Mapping (SPM). Voxel based morphometry (VBM)[48] is a neuroimaging analysis technique that allows investigation of focal differences in the whole brain in an unbiased way, using the statistical approach such as Statistical Parametric Mapping (SPM) [49]. A typical measure of structural difference among populations is derived from a comparison of the local composition of different brain tissue types (e.g. GM, WM, CSF etc). VBM has been known to be sensitive to these differences, while discounting positional and other large scale volumetric differences in gross anatomy. The output of VBM is an SPM showing regions where GM concentration differs significantly among the groups.

The VBM technique typically uses T1-weighted volumetric MRI scans due to a better characterization of the tissue resolutions and then applies a series of t tests at every voxel in the image. The most essential step for VBM is to register every brain image with a template, which is constructed by registering, smoothing, and averaging of all the images in the control group. VBM has been a very popular method in the medical image analysis application because of its usefulness in characterizing subtle changes in brain structure in a variety of diseases. Several methods have also been proposed to detect FCD lesions based on VBM[34]. However, there are several limitations of VBM based methods [50] that makes this method not suitable for the study on FCD lesion detection, which we illustrate as follows:

1. *Normalization and Segmentation Confounds* VBM based method is sensitive to mis-registrations, and the registration task can be more challenging

with the presence of disease. Also, the registration can be subject to the size difference between brains, especially when dealing with children subject. Moreover, image quality and artifact can also greatly affect the performance VBM based methods. In our study, images are acquired from different scanners with various image resolutions and subjects from different age group, it is likely that significant effects might be attributable to the differences between scanners and normalization error rather than to the subtle FCD lesions, therefore this study may not be suitable for VBM based analysis.

2. *Pre-processing of atypical brains* The investigation of brains containing abnormal pathologies may also be problematic, due to the difficulties when applying spatially normalization and segmenting to atypical brains. For example, the pre-processing steps in VBM requires segmentation of GM, WM and CSF, which can be challenging for subjects with FCD, where the GM and WM are often hard to define in the lesional area.

*Discriminative feature-based analysis* tries to detect lesions based on a set of discriminative features and a decision rule learned from previous examples. This type of analysis extracts characteristics to differentiate between healthy and normal regions. The performance is highly dependent on the selected data representations, or *features*, while selecting a good set of features requires domain knowledge and is highly application-specific. Also, features are often calculated from a small subsets of the brain. Dividing the brain into subsets could potentially lose valuable spatial structural information that could serve as useful information for making diagnostic decision. The problem become more prominent especially when features are calculated voxel-wise. The idea of using discriminative features for the detection of FCD lesions was first proposed by Bernasconi et al. [37]. Three computational features: cortical thickness, gradient magnitude, and relative intensity, are computed for each voxel in the brain to capture the three common characteristics of FCD lesions on T1 weighted MRI. The features were refined [36] and combined with volume-based textural features in a later study [7],

we call this set of features the Antel's features. In several follow-up works, the Antel's features were quantitatively analyzed [43][3] and used as the initialization step for the method of [41] focused on delineation. While this set of features is proven to be highly related to lesional characteristics [43] and reported to detect lesions successfully, the issue of false positives and low coverage rates [7] are its major limitations. More detail on Antel's features are presented in the following.

### Antel's Features

Antel's features [36] and its associated CAD framework [7] are the most widely referenced methods in the literature for automatic FCD detection and serve as the foundation of many later research, including this thesis. In this section, we discuss the methodology, usefulness, and limitations and the Antel's features.

The Antel's features consist of three voxel-based attributes: thickness, intensity and absolute gradient. The calculations of these features are illustrated as follows.

- *Cortical Thickness*

  The cortical thickness measurement used by Antel et al [36] is developed by Jones et al [51]. This method calculates the cortical thickness as an electrostatic field, $\Psi$ with the requirement that the inner and outer cortical surfaces are predefined. $\Psi$ is obtained as a solution to the Laplace's Equations and can be written as:

  $$\nabla^2\Psi = \partial^2\Psi/\partial x^2 + \partial^2\Psi/\partial y^2 + \partial^2\Psi/\partial z^2 = 0 \tag{2.9}$$

  When solve for $\Psi$, a unit vector field can be computed over the cortex, which is,

  $$N = -\nabla\Psi/\| - \nabla\Psi\| \tag{2.10}$$

  The cortical thickness is computed by starting at any point at one of the surfaces and the integration is along the vector $N$. The technique requires a

hard segmentation of the MRI brain volume into GM, WM, and CSF, which is a major limitation which we will describe in the following section.

- *Relative Intensity*
  The relative intensity attribute $I_R$ is defined to detect hyperintense T1 signal within the GM. It is defined as

$$I_R = 1 - |g - B_g|/B_g. \tag{2.11}$$

  where $g$ is the intensity value at a given voxel and $B_g$ is the threshold intensity of the GM and WM peaks as determined from the gray-level histogram of the MRI volume.

- *Absolute Gradient*
  The gradient magnitude for Antel's features is the same as the standard $1^{st}$ order partial derivative. The gradient magnitude was calculated by first going through convolution of the MRI volume with three-dimensional Gaussian kernel (FWHM=3mm) in the Fourier domain. The blurring of the GMWM was modeled with the gradient magnitude calculated in the spatial domain over a cubic neighborhood.

**Limitations of Antel's Approach**

Antel proposed using three computational operators which directly link to three important characteristics of FCD: increased cortical thickness, blurriness in the GM/WM junction and hyper-intensified T1 signal. The features provide valuable quantitative measurements to assist diagnosis of FCD. However, there are some limitations in Antel's features which can be summarized as follows:

1. *Calculation of the Cortical Thickness* The approach selected by Antel et al [51] to measure cortical thickness is based on the Laplace equation ($\nabla^2\Psi = 0$) from mathematical physics. The assumption of this method is that the cortex of each hemisphere is a volume bounded by two surfaces, $S$ and $S'$,

which is defined by the GM-CSF and GM-WM surface. These surfaces can be hard to determine, especially in the case for subject with white matter lesions like FCD. Defining such surfaces before the estimation of cortical thickness can be making a decision before the estimation.

2. *Reliability of Gradient Magnitude* Gradient magnitude is widely used as an image processing operator to extract information from image to quantify variation (e.g. edge detection). In Antel's work [36], gradient magnitude is used to capture the characteristic of blurriness. However, low gradient magnitude typically implies lack of information or variation, and does not necessary imply blurriness. Therefore, using gradient magnitude can be misleading. For example, the lowest gradient magnitude often appear at voxels in areas consisting of only WM or GM, which is not typically the lesional area (GM/WM boundary) of interest.

3. *Definition of Relative Intensity* The definition of relative intensity used by Antel et al in eq.(2.11) has several limitations. First, the value of this feature is determined by the absolute intensity value of a given voxel relative to a pre-defined threshold $B_g$. From our observations of the clinical data of FCD, the lesional voxels do not necessary have a hyper- or hypo-intensified signal, but most of the time, resemble the intensity of GM. One example can be seen in Fig 2.1, there are two boxed areas in a given subject. The top box represents FCD lesions and the lower box contains a typical GM area. One can see that the distributions of the intensities from the two boxed area are very similar. Therefore, it demonstrates that using relative intensity as a feature has little discrimination ability to separate between GM area and FCD lesions.

4. *Voxel-based Analysis* One of the major limitations of Antel's features is the use of voxel-based analysis. It is well-established in the image processing and pattern recognition literature that per-pixel or per-voxel classifications do not make full use of valuable spatial information in the image [52]. For

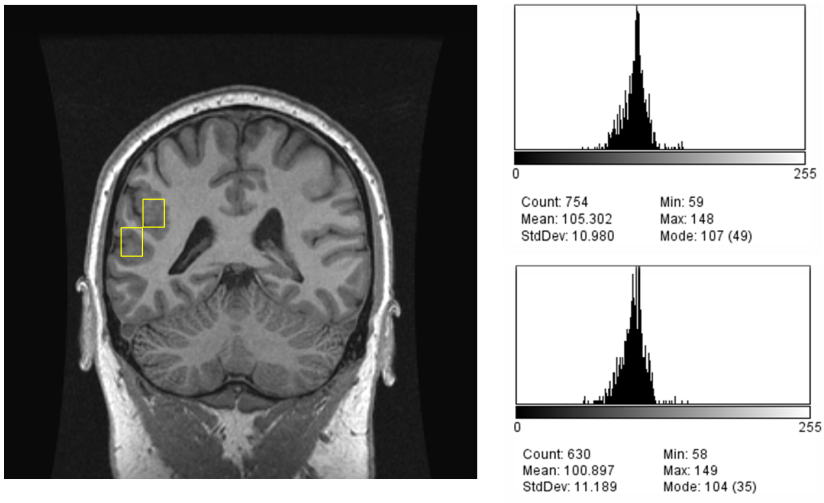


Figure 2.1: Comparison of voxel intensities from a FCD lesional area (top box) and a typical GM area (bottom box)

medical image analysis and CAD applications, target lesions tend to cover several voxels in the surrounding neighborhood. One voxel being classified as lesional should make its surrounding tissues more likely to be lesional as well. A classification method that fails to incorporate spatial information has its limitations.

### 2.5.2 Hyperintense lesion segmentation on T2-weighted MRI

T2-weighted Fluid Attenuated Inversion Recovery (FLAIR) image sequences are commonly used for brain imaging because they can suppress signal strength for cerebrospinal fluid (CSF) so as to bring out the hyperintense lesions, such as multiple sclerosis (MS)[53] or FCD in our study. Detection of such lesions is less challenging, however, segmentation of such lesions is often required. Manual segmentation is time consuming and often suffers from inter- and intra-rater variability [54]. The segmentation of lesions in MR images can also provide quantitative measurements, which may be important for monitoring of disease, drug

treatment evaluations, and provide aid to surgery and treatment planning [55].

A variety of approaches have been proposed, both for image segmentation in general (see survey [56]) and medical image segmentation in particular (see survey [55]). Each segmentation algorithm is highly dependent on specific applications, imaging modality, and other parameters. There is currently no single segmentation algorithm that provides successful results for every clinical application. Also, it is hard for one to give a complete review of the segmentation literature. Therefore, we restrict our discussion on the applications of hyperintense lesion segmentation on T2-weighted MRI images. Based on the summary from [55], the medical image segmentation can be divided into five categories:

1. thresholding approaches [22][57][58]
2. region growing approaches [59][60]
3. pixel classification methods [61][62]
4. deformable models [63]
5. atlas guided approaches [64]

We will review the literatures of hyperintense lesion segmentation on T2 MRI images based on these categories. For *thresholding approaches*, segmentation is achieved by generating binary partitions of the image intensities. The thresholding approaches require determination of an intensity value, called the *threshold*, that separates the two classes. Mitchell [22] and Filippi [57] apply thresholding in the detection of hyperintense brain lesions. However, the former requires the user to select the threshold manually based on the histogram while the latter requires manual selection of the Region Of Interest (ROI).

*Region growing methods* extracting regions of the images that are connected based on predefined criteria. This criteria can be based on intensity information, textures, colors, or edges in the image. The most common type of region growing method involves the user selection of seed points. The initial region begins as the

exact locations of these seeds. Warfield [59] and Pan [60] proposed region growing based methods. However, region growing methods suffer from the difficulty in selecting the seed points and determining the stopping criterion.

*Pixel classification methods* can be further divided into several sub-categories: classifiers, clustering approaches, Markov random field models, and artificial neural network. Instead of working on intensity or the image space alone, these type of methods operate on a user-defined feature space. In a pattern recognition problem, a feature space is an abstract space where each pixel/voxel is represented as a $n$-dimensional feature vector. In the applications of hyperintense lesion segmentation, intensity-based classification (supervised), clustering approaches (unsupervised), and Markov random field modeling are often chosen. Here we summarize some of the key research works that utilizes these three approaches.

- *Intensity-based clustering approaches*
  Clustering algorithms, or *unsupervised* algorithms, are defined as algorithms that require no previously provided class labels. Three main clustering algorithms are k-means, fuzzy c-mean [65], and expectation maximization (EM) algorithm. These types of methods are frequently applied for medical image segmentation. Although clustering algorithms do not require the training of data, they also do not directly incorporate spatial information, and can therefore be sensitive to noise and intensity variations [55]. The robustness of the intensity-based clustering approach can be improved by adding spatial information such as Markov Random Field modeling which is described later in the Markov Random Field section.

- *Intensity-based classification approaches* Classifier methods or *supervised* algorithms are pattern recognition techniques that seek to partition of a feature space derived from the image using data with known labels. The simplest example for a non-parametric classifier is the k-nearest neighbor (kNN). Another widely-used parametric classifier is the maximum likelihood (ML)

or Bayes classifier. The classifier assumes the pixel is drawn from a mixture of probability distributions, say Gaussians. The classifier is learnt by optimizing the parameters associated with the distributions. Like the clustering methods, the major weakness of this type of methods is the lack of incorporation of spatial information.

- *Markov random field model* Markov random field (MRF) models spatial interactions within a spatial neighborhood. MRF modeling is appealing in the computer vision and pattern recognition for several reasons [66]:(1) It allows one to systematically develop algorithms based on sound principles (2) ease of deriving quantitative performance measures (3) MRF can be used to incorporate various prior contextual information in a quantitative way. Many works have also been proposed to use Markov Random Field modeling specifically for detecting hyper-intensity signal in the medical image processing domain [67][68].

In this section, we review previous methods on segmentation of hyperintense lesions in general, including the MRF approach which is used to incorporate spatial information with the intensity information. More details of the MRF approach we used is described in Section 4.2. In the next section, we review the method designed particularly for segmentation of FCD lesions on T2 weighted FLAIR MRI .

### 2.5.3 FCD Lesion detection and segmentation on T2 weighted MRI

Compared with the well-established T1-weighted FCD automated detection algorithms, fewer investigation [69] have been focused on FCD detection on T2 FLAIR MRI due to the lower image resolution that can be achieved using the T2-weighted MRI at lower magnetic field. However, through the development of imaging technology, research has shown that MR imaging at 3 Tesla and the use of

phased arrays have a higher signal-to-noise ratio compared to traditional 1.5 Tesla imaging, and has the potential to increase the accuracy of FCD diagnosis process [70] (see Fig. 2.2). On the example of Fig. 2.2, the FCD lesion is much easier to detect on the T2 weighted MRI images as opposed to T1 weighted MRI. In 2008, Focke et al. [69] proposed a whole brain voxel-based technique the utilizes Statistical Parametric Mapping (SPM) using the normalized FLAIR signal intensity that compares 25 healthy controlled subjects and 25 patients with neurologists reported FCD. The method reports 88% identification of the FCD lesions with only one false positive in the control group.

This method is the first work that is proposed to detect FCD lesion on T2 FLAIR sequences. However, there are several limitations which allow room for improvements for this approach, which we summarize as follows:

1. *Pre-processing* Smoothing and normalization play a crucial role in the accuracy of the proposed method. The approach requires the images of multiple subjects to be registered together by some form of spatial normalization, which can be challenging with varieties in the data set.

2. *Semi-automatic* The method requires manual threshold determination, which requires expert knowledge and can often be time consuming.

3. *Intensity based feature* The use of only intensity based feature implies inevitable number of false positives due to image noise and inhomogeneity.

4. *Lack of information from T1 images* Information from T1-weighted MRI is not combined in assisting the diagnosis.

### 2.5.4 Summary

In this section, we reviewed previous approaches for FCD lesion detection and segmentation on both T1 (Section 2.5.1) and T2 (Section 2.5.3) weighted MRI. The pros and cons of these approaches were discussed, and thus far, no single

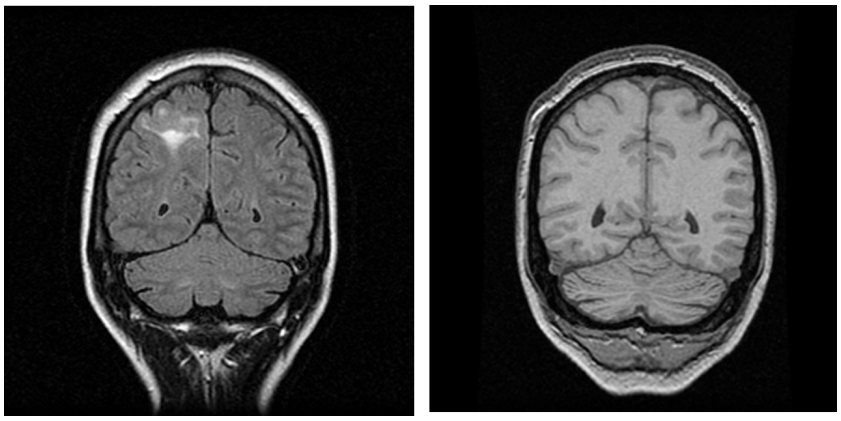

Figure 2.2: FCD lesions on T2 weighted MRI(left) and T1 weighted MRI(right)

method can solve the problem without assumptions and limitations. In the following chapters, we illustrate our proposed methods to approach the FCD lesion detection and segmentation problem on both T1 and T2 weighted MRI in order to provide improvement on the existing strategies.

# Chapter 3

# Automated FCD Lesion Detection on T1 weighted MRI

In this chapter, the proposed CAD system of FCD lesion detection on T1 weighted MRI is described. In Section 3.2, we introduce the proposed volume-based features. The analysis using Naive Bayes Modeling of the proposed features is presented in Section 3.3. In Section 3.4, cluster-wise analysis is used to reduce the number of false positives from the previous stage. The classification of the clusters using Support Vector Machine is included in Section 3.5. Finally, we summarized the methods proposed for FCD lesion detection on T1 weighted MRI in Section 3.6.

## 3.1 Pre-Processing

Each subject's volumetric MRI data first undergoes several pre-processing steps, which includes skull-striping, removal of non-brain tissue, normalization. After such preprocessing, the data is normalized into the same $256 \times 256 \times 256$ size with intensity ranges from 0 to 255. All of the mentioned preprocessing steps are

achieved by the software package FreeSurfer [71][72][73]. An example is presented in Fig 3.1.

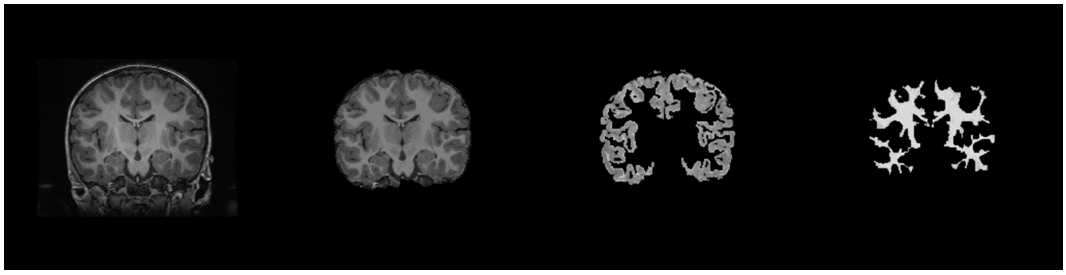

Figure 3.1: Preprocessing results (left to right): Original brain, brain-extraction, GM and WM tissue classification from FreeSurfer

## 3.2 Volume-based Feature Extractions

In this thesis, contrast to the traditional voxel-based features and analysis, a volume-based features and analysis are used. The volume-based approach allows ones to study the distributions of the features within a spatial neighborhood, which can not be done with the traditional voxel-based approach.

A subject's $256 \times 256 \times 256$ preprocessed image volume data is divided into overlapping $8 \times 8 \times 8$ volumetric cubes. Each cube is shifted by one voxel to avoid boundary effects, and to ensure that the lesion will be entirely covered. Seven features are computed for each cube. These features were selected to more fully represent the distributional characteristics of the discriminative attributes of the cortical thickness and gradient vectors. The features are : cortical thickness, the skewness of cortical thickness, the kurtosis of cortical thickness, absolute gradient, and the variance/skewness/kurtosis of the gradient vectors' orientations. Details on the calculation and the attributes captured by the features are as follows:

*i. Cortical thickness* The method used for the calculation of cortical thickness in Antel's work [36] belongs to the category of Laplace methods [51]. These type of methods solve for Laplaces equation in the GM region with the boundary condition on each of the two surfaces. However, in the study of FCD lesions, the GM and WM junctions are undefined, the requirement for presegmentation of the two surface greatly reduces the accuracy of this technique.

Cortical thickness is calculated by the method of Aganj et al.[74]. The thickness of the GM at a given voxel was defined as the minimum line integral of the probability map of the GM over all lines that pass through the voxel.

$$T(\overrightarrow{x}) = \min_{l \in L_{\overrightarrow{x}}} \int_l P_{GM}(r) dl, \tag{3.1}$$

where $T(\overrightarrow{x})$ is the thickness of the GM at a given point $\overrightarrow{x} = (x, y, z)$. $P_{GM}(\overrightarrow{x})$ is the probability of the point $\overrightarrow{x}$ belonging to GM and $L_{\overrightarrow{x}}$ is the set of all lines in three-dimensional space. The thickness feature for each cube is the average thickness of the thickness measure for each voxel. The main advantage of choosing this method for the calculation of cortical thickness is the fact that the method requires no hard labeling of the tissues. The Laplace method [75] used by [36] requires a presegmentation of the GM and WM surface, which is unsuitable for our problem with undefined GM and WM junctions.

To compute $P_{GM}$, a Gaussian Mixture Model (GMM) [76] is used to fit 3 Gaussian curves to the brain intensities (1 for cerebral spinal fluid (CSF), 1 for GM, 1 for WM) using the Expectation-Maximization (EM) algorithm. The brain tissue can be modeled as $P(\overrightarrow{x}) = \alpha_c P_{CSF}(\overrightarrow{x}) + \alpha_g P_{GM}(\overrightarrow{x}) + \alpha_w P_{WM}(\overrightarrow{x})$, where each $\alpha$ is the coefficient for each Gaussian component. Using the second Gaussian component (with second largest mean) $P_{GM}$, we modeled the probability of each voxel being GM as

$$P_{GM}(\overrightarrow{x}) = \frac{1}{\sqrt{2\pi\sigma_g^2}} \exp\left(\frac{-(I(\overrightarrow{x}) - \mu_g)^2}{2\sigma_g^2}\right), \tag{3.2}$$

where $I(\overrightarrow{x})$ is the intensity value of the point $\overrightarrow{x}$, and $\mu_g$ and $\sigma_g$ are the mean and standard deviation of the second component from the GMM, respectively. One example of the calculated thickness measurement can be seen in Fig 3.2, one can see that the cortex around lesional area is measured with higher thickness value as indicated as higher intensity in the thickness map.

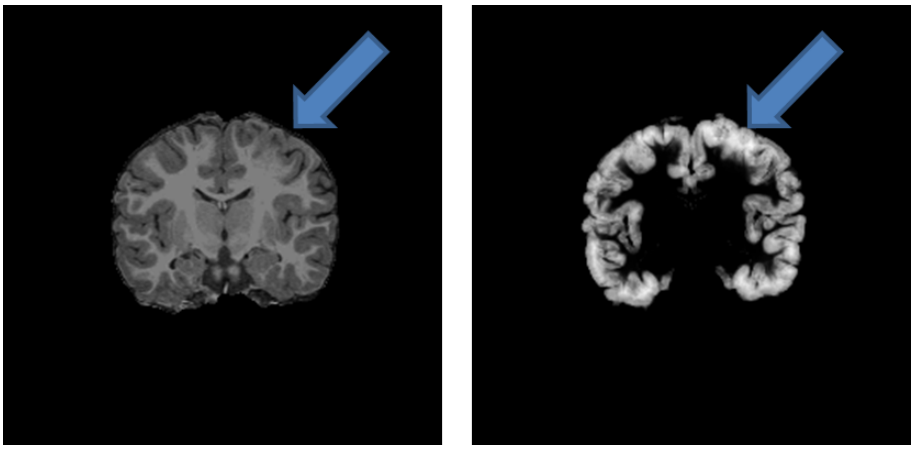

Figure 3.2: The original T1 weighted MRI image (left) and the corresponding thickness map (right). The lesional area is marked with arrows.

*ii. Skewness and kurtosis of the cortical thickness*

Skewness is defined as the third standardized moment of a random variable, which is a measure of asymmetry of the data around the sample mean. Here we calculate the sample skewness of the thickness measures in a given cube. For the thickness data $t_1, t_2, ..., t_N$ in the $ith$ cube, the sample skewness can be calculated as,

$$skewness_t(i) = \frac{\sum_{n=1}^{N}(t_n - \bar{t})^3}{(N-1)\sigma_t^3} \tag{3.3}$$

In a similar manner, the fourth standardized moment, kurtosis, is also calculated for the thickness data.

$$kurtosis_t(i) = \frac{\sum_{n=1}^{N}(t_n - \bar{t})^4}{(N-1)\sigma_t^4}, \tag{3.4}$$

where $\bar{t}$ is the mean of the measured thickness, $\sigma_t$ is the standard deviation, and $N$ is the number of voxels in a cube, which is 512 in this study.

*iii. Absolute gradient*

Gradient is a good indicator of the amount of intensity transitions of voxels and are often used for detecting edges or measuring sharpness. A cube with larger average gradient value, which often associates with the presence of sharp edges is less likely to contain lesions of FCD. For a 3D image volume $f$, the 3D gradient at a given voxel point is calculated as $\nabla f(x,y,z) = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}) = (G_x, G_y, G_z)$. Then for each cube, the gradient feature is obtained as

$$gradient(i) = \sum_{x,y,z \in cube(i)} \sqrt{|G_x|^2 + |G_y|^2 + |G_z|^2} \tag{3.5}$$

However, low gradient value does not necessary imply blurriness. Therefore we proposed the next feature to work as an additional indicator for blurriness.

*iv. Statistics of the orientations of gradient vectors*

For most of the cubes that contain the boundary of GM and WM, it is observed that normal tissues often contain clear edges. A clear edge would have the directions of the gradient vectors consistently pointing normal to the edges. On the other hand, lesional cubes with a blurred WM/GM junction will certainly not have gradient vectors that point in a clear dominated direction (See Fig. 2). In image quality literature, non-reference blurriness metric can be measured using the orientation of the gradient vectors [77]. Inspired by this fact, the variance of the angle of the gradient vectors in a given cube is chosen as one of our

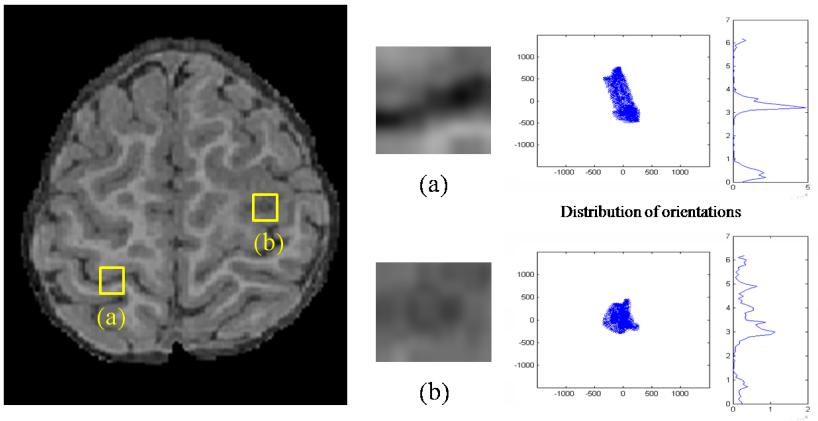


Figure 3.3: The distribution of orientations of gradient vectors from two different regions of the brain. (a) normal (b) with FCD lesions

features. For every fixed $z$, the 2D gradient calculation for the $x-y$ plane is $\nabla f(x,y) = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}) = [G_x, G_y]$. And the direction can be calculated as,

$$d(x,y) = atan2(G_y, G_x) \tag{3.6}$$

The value is in the range of $[-\pi, \pi]$ Similarly, the directions of $d(x,z)$ and $d(y,z)$ are also calculated. For every $x-y$ plane in cube $i$, we calculated the directions and the variance can be obtained by,

$$var_{xy}(i) = \sum_{x,y \in cube(i)} (d(x,y) - \overline{d(x,y)})^2, \tag{3.7}$$

where $\overline{d(x,y)}$ is the average gradient directions for all the $x-y$ planes in a given cube. Then for each cube, the average variance of gradient directions in three-dimension is,

$$var_g(i) = \frac{var_{xy}(i) + var_{xz}(i) + var_{yz}(i)}{3}. \tag{3.8}$$

Similar to the way we computed the variance calculation in Eq.(3.7) and Eq.(3.8), the skewness for the orientations of gradient vectors is calculated for the three respective 2D planes and then averaged. The skewness for the directions of gradient

vectors is,

$$skew_g(i) = \frac{skew_{xy}(i) + skew_{xz}(i) + skew_{yz}(i)}{3}. \tag{3.9}$$

Similarly, the kurtosis is computed in as Eq.(3.9).

$$kurt_g(i) = \frac{kurt_{xy}(i) + kurt_{xz}(i) + kurt_{yz}(i)}{3}. \tag{3.10}$$

The seven features are able to provide statistics of cortical thickness and the gradient vectors of a certain volume region of the brain. The volume-based features are calculated by dividing the volume of T1-weighted MRI into $8 \times 8 \times 8 \times 8$ volumetric cubes. The statistics are calculated based on the feature distributions formed by the $N = 512$ voxels in the cubes. Assume for each voxel $v_i| \in [1, N]$ in a given cube, there is a cortical thickness measurement $t_i$ and gradient vector measurement $\hat{g}_i$ with the magnitude $|\hat{g}_i|$ and angle $\angle\hat{g}_i = \psi_i$. The seven features and their descriptions are summarized in Table 3.1.

Table 3.1: The seven volume-based absolute features

| Feature | Feature description |
|---|---|
| Thickness | Average of $t_i$ |
| Thickness skewness | Skewness of distribution of $t_i$ |
| Thickness kurtosis | Kurtosis of distribution of $t_i$ |
| Absolute gradient | Average of $\|\hat{g}_i\|$ |
| Orientation variance | Variance of the distribution of $\psi_i$ |
| Orientation skewness | Skewness of the distribution of $\psi_i$ |
| Orientation kurtosis | Kurtosis of the distribution of $\psi_i$ |

## 3.3 Volume-based feature classification using Naive Bayes Modeling

After the calculation of the volume-based features, a Naive Bayes framework which is introduced in Section 3.3 is used for the modeling and classification. In this

study, 2 classes are of interest $\{H, L\}$, which $H$ and $L$ stands for healthy and lesional respectively. The attributes are the proposed volume-based features $f_n$ described in Section 3.4. With the Naive Bayes prediction rules, to learn whether a given volume is lesional or not, we seek a target function $f : X \mapsto Y$. In a Bayesian framework, to learn this function, we use the training data to estimate the class conditional probability $P(X|Y)$ and class prior $P(Y)$.

*($\alpha$) Class prior*

If we assume the number of classes $C$ is small, which is true in our case with $C = 2$, then we can estimate the class prior $P(Y)$ by treating $Y$ as a multimodal random variable:

$$P(Y = y_k|\pi) = \pi_k, \tag{3.11}$$

where $\pi$ is a vector of class probability. The maximum likelihood estimation (MLE) is simply

$$\hat{\pi}_k = \frac{\sum_j I(y_j = y_k)}{N} = \frac{N_k}{N}, \tag{3.12}$$

where $N_k$ is the number of training instances that belong to class label $k$.

*($\beta$) Class conditional probability*

The main issue is how to estimate the class condtional probability $P(X|Y)$. With the naiveness assumption, for any given $y_k$, the probability can be calculated using eq(2.2). A common approach is assume $P(X_i|Y = y_k)$ follows a Gaussian distribution:

$$P(X_i = x|Y = y_k) = \frac{1}{\sqrt{2\pi\sigma_{ik}^2}} \exp -\frac{1}{2}(\frac{x - \mu_{ik}}{\sigma_{ik}})^2. \tag{3.13}$$

By using the Gaussian assumption, the problem of estimating the probability distribution is reduced to estimating the mean parameter $\mu_{ik}$ and variance parameter $\sigma_{ik}$ using MLE as follows:

$$\hat{\mu}_{ik} = \frac{1}{\sum_j \delta(Y^j = y_k)} \sum_j X_i^j \delta(Y^j = y_k). \tag{3.14}$$

$$\hat{\sigma}_{ik}^2 = \frac{1}{\sum_j \delta(Y^j = y_k)} \sum_j (X_i^j - \hat{\mu}_{ik})^2 \delta(Y^j = y_k). \tag{3.15}$$

## 3.4 Cluster-based Features Extraction

In addition to the volume-based features, we designed another set of features based on *clusters*. In this study, the clusters are defined as a spatial neighborhood collections of $8 \times 8 \times 8$ volumetric cubes. The calculated volume-based features within a cluster should have a certain degree of spatial homogeneity. The purposes of using the cluster-based feature are: (1) study of feature distributions within a spatial neighborhood (2) use cluster-wise differential information to assist further classification. In the following analysis, we calculate the difference between feature distributions of the target cluster area and other parts of the brain. These kinds of features, instead of evaluating the absolute feature distributions, we are more intrusted in the *differences* between features extracted from different features. To highlight this concept, these features are called the *differential features* in this thesis.

The cluster-based differential features are based on the differences of the absolute volume-based features from three different areas inside the brain. Here we define the selection and feature calculations of the regions.

1. *Target cluster* The target cluster is from the voxel clusters that are marked as positive (lesional) from the previous volume-based feature classification using a Naive Bayes classifier as we described in Section 3.3. These positive areas generally form in clusters instead of random scattering of isolated voxels because that voxels in the vincinity of each other often share similar characteristics . In this study, clusters are defined as any disjoint areas: any area with no connection to other positive output voxels from an individual cluster. An example can be seen in Fig 3.4. In Fig 3.4, there are three voxel clusters presented; however, only cluster 3 is associated with the

actual lesional area. The other two clusters are false positives (FPs). In words, clusters that are predicted as lesional from the previous stage serve as the candidate or target clusters for further analysis as we describe below.

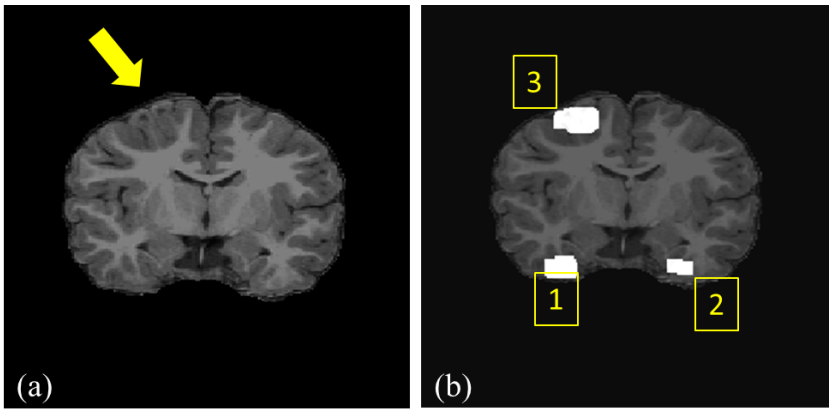


Figure 3.4: Example of positive clusters.(a) Original image with actual lesion indicated by arrow (b) The positive voxels highlighted from volume-based feature classification using Naive Bayes Classifier. Only cluster 3 is associated with the actual lesion

2. *Neighborhood area of the target cluster* In order to compare a candidate lesional target clusters with its surroundings, for every cluster $C_j$ in a given subject, we define its surrounding neighborhood $S_j$ as the minimum rectangular volume that covers $C_j$ and excludes all voxels that are contained in $C_j$. The definition of the neighborhood area help us understand how much the target cluster is different from its immediate surrounding voxels, which can possibly provide useful information for the diagnosis of FCD lesions.

3. *GM/WM boundaries area* FCD lesions occur mostly on the GM and WM junctions. Therefore, it is desirable to compare candidate lesions with other boundaries of GM and WM. In theory, the features of an actual lesional area should have very different feature distributions with other normal GM/WM boundaries. To include all the GM/WM boundaries, for each subject, the

brain volume is again divided by $8 \times 8 \times 8$ subsets. We define a volume cube is at the boundary of GM and WM if more than 1/3 of the voxels in the cube is WM, and also more than 1/3 of the voxels in the cube is GM. (see Fig 3.5) Here the definition of GM and WM is from the Gaussian Mixture Model defined in [78]. Assume all the cubes volume at GM/WM boundaries is $V$, we can then calculate sample distributions or histograms for all thickness measurements, gradient magnitudes, and gradient orientations for $V$. Then we compute the histogram distances between the GM/WM area and all of the lesional candidate clusters using also the Chi-square distances and build a feature vector for each cluster. More details on building the cluster-based feature vector is defined in the next section.

### 3.4.1 Building the cluster-based feature vector

For each cluster-based feature vector, it is consisted of a total of 13 attributes. The first 7 feature vector attributes are calculated in a similar way as the volume-based features described in Section 3.2. The only difference is that the statistics of thickness, $t_i$, and gradient vectors, $\overrightarrow{g}_i$, are calculated over measurements of the entire cluster $C_j$ and $S_j$ respectively, instead of individual volume cube as was done in Section 3.2. Then we represent all the measurements for each cluster $t_i$, $|\overrightarrow{g}_i|$ and $\phi_i$ into histograms.

Let us denote the thickness measurements in the candidate lesion locations $C_j$ as $t_i^{c_j}$ and in the surrounding locations $S_j$ as $t_i^{s_j}$. Similarly, the gradient measurements are presented as $\overrightarrow{g}_i^{c_j}$ and $\overrightarrow{g}_i^{s_j}$. $h(t_i^{c_j})_M$ is the histogram representations for all the $t_i$'s in cluster $C_j$ in a total of $M$ bins. The distance of the two histograms between lesional candidate and its surroundings, $h(t_i^{c_j})_M$ and $h(t_i^{s_j})_M$ is calculated using a symmetrized approximation of $\chi^2$ distance as,

$$d(h(t_i^{c_j})_M, h(t_i^{s_j})_M) = \sum_{m=1}^{M} \frac{(h(t_i^{c_j})_m - h(t_i^{s_j})_m)^2}{h(t_i^{c_j})_m + h(t_i^{s_j})_m}, \tag{3.16}$$

where index $m$ represents the $m$-th bin in the histograms. Note that the histograms are shifted to the same range and are normalized as $\sum_{m=1}^{M} h(t_i^{c_j})_M = \sum_{m=1}^{M} h(t_i^{s_j})_M = 1$ before calculating the distances. The calculations also apply to the histogram distances of gradients.

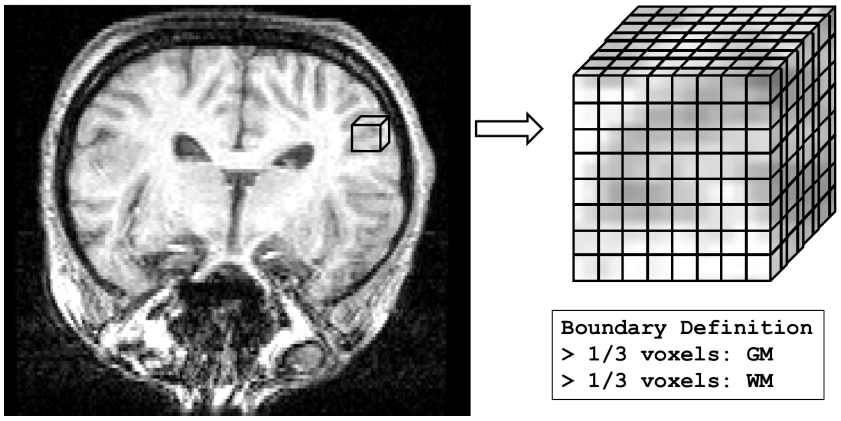


Figure 3.5: Definition of GM/WM boundary is any $8 \times 8 \times 8$ volume that contains more than 1/3 GM voxels and more than 1/3 WM voxels

The calculation of cluster-based differential features is summarized in Table 3.2. To summarize the building of the cluster-based feature, we outline the analysis in the following steps:

**Step 1:** Identify the number of target clusters. A target cluster is defined as a spatial neighborhood of $8 \times 8 \times 8$ voxel cubes that are marked as "lesional" from the previous volume-based features and Naive Bayes classifier. The positively marked voxel cubes should be directly connected to at least one ther voxel cube in the same cluster.

**Step 2:** For each separate cluster, the histograms of the seven volume-based features is built. Except for the target cluster, the feature histograms from the neighborhood, and also the entire GM/WM boundaries are also calculated.

**Step 3:** The cluster-based feature vector is built by averaging the seven absolute

volume-based feature in Table 3.1 and including the six differential features in Table 3.2

After finishing building these cluster-based features, a classifier should be selected. Unlike our previous voxel-based analysis, prior expert knowledge is not available here. The relationships of the differential features from target clusters, its neighbors and other GM/WM are unknown. Therefore, in this case, a discriminative classifier Support Vector Machine (SVM) becomes appropriate and is used in this thesis. The clusters are classified using a SVM classifier with class labels of either healthy (0) or lesional (1). In the next section, the formulation and parameter selection for our use of SVM is briefly described.

Table 3.2: The cluster-based differential features

| Differential Feature | Feature Description | Expression |
|---|---|---|
| Thickness (neighbor) | $t_i$ distribution difference between $C_j$ and $S_j$ | $d(h(t_i^{c_j}), h(t_i^{s_j}))$ |
| Absolute gradient (neighbor) | $\lvert\overrightarrow{g}_i\rvert$ distribution difference between $C_j$ and $S_j$ | $d(h(g_i^{c_j}), h(g_i^{s_j}))$ |
| Gradient orientation (neighbor) | $\phi_i$ distribution difference between $C_j$ and $S_j$ | $d(h(\phi_i^{c_j}), h(\phi_i^{s_j}))$ |
| Thickness (GM/WM) | $t_i$ distribution difference between $C_j$ and $V$ | $d(h(t_i^{c_j}), h(t_i^{v}))$ |
| Absolute gradient (GM/WM) | $\lvert\overrightarrow{g}_i\rvert$ distribution difference between $C_j$ and $V$ | $d(h(g_i^{c_j}), h(g_i^{v})$ |
| Gradient orientation (GM/WM) | $\phi_i$ distribution difference between $C_j$ and $V$ | $d(h(\phi_i^{c_j}), h(\phi_i^{v}))$ |

## 3.5 Cluster-based feature classification using Support Vector Machines

From Section 2.4.2, we know that the discriminant function of the SVM with linear classifier in the dual form can be defined as:

$$f(x) = sign(\sum_{i,j=1}^{m} a_i y_i (x_i^T x_j) + b), \tag{3.17}$$

where $a_i$ is the weight coefficient, $b$ is the offset term,$y_i$ is the prediction and $x_j$ and $x_i$ are the data vector as defined in Section 2.4.2. In the case of non-linearly separable classes, the kernel approach is then used. SVM classifier first maps the input vector $x_i$ into a higher order feature space using the kernel transformation:

$$x^T x_i \longmapsto \phi^T(x_j).\phi(x_i). \tag{3.18}$$

The dot product of $\phi^T(x).\phi(x_i)$ is achieved by using a kernel function $k(x_i, x_j)$. Equation (3.18) can be re-written as,

$$f(x) = sign(\sum_{i,j=1}^{m} a_i y_i k(x_i, x_j) + b). \tag{3.19}$$

There are a number of kernels available as introduced in Section 2.4.2. In this study, we select the cubic polynomial kernel. The reason for us to make the selection is that from our preliminary experiments, it provides better discrimination than the $1^{st}$, $2^{nd}$ order polynomial kernel, and the Radial basis function kernel. The discriminant function of a SVM classifier with the third degree polynomial kernel can be written as,

$$f(x) = \sum_{i=1}^{d} a_{iii} x_i^3 + \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} a_{iij} x_i^2 x_j + \sum_{i=1}^{d-1} \sum_{j=i+1}^{d}$$
$$a_{ijj} x_i x_j^2 + \sum_{i=1}^{d-2} \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} a_{iik} x_i x_j x_k + \sum_{i=1}^{d} a_{ii} x_i^2 + \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} a_{ij} x_i x_j + \sum_{i=1}^{d} a_i x_i - b, \tag{3.20}$$

where $d$ is the number of features, $a$'s are weight coefficients, $b$ is the offset term and $x_j$ are the data vector as defined in Section 2.4.2. The effectiveness of SVM

depends on the choice of kernels, kernel parameters, and soft margin parameter $C$. The coefficient $C$ affects the trade-off between complexity and proportion of non-separable instances and must be chosen by the user. Typically, each combination of parameter choices is checked using cross validation, and the parameters with best cross-validation accuracy are picked. The final model, which is used for testing and for classifying new data, is then trained on the whole training set using the selected parameters. In this thesis, we also used a cross-validation technique to determine the accuracy of SVM, more details on the training and validation is described in Section 6.2.1.

## 3.6 Summary

In this Chapter, we provided detail information on the proposed CAD system for FCD lesion detection on T1-weighted MRI. The proposed procedures for the CAD system is summarized in Fig 3.6. A two-stage classification system is used in this work. In the first classification stage, a generative Naive Bayes classifier is used for volume-based analysis. After the preliminary volume-based classification, lesional clusters which are marked by the first classifier serve as input for the second cluster-based classification system. In the second cluster-based classification, a new set of differential features are used and a more sophisticated, discriminative Support Vector Machine (SVM) classifier with a third-order polynomial kernel is used for cluster-based classification. We find it reasonable to first use a less computationally expensive, generative classifier like Naive Bayes to provide overall characteristics of all of the tissues in the brain, which requires a large amount of calculations. From the first stage, we eliminate a large number of tissues from consideration of potential lesions, however, the remaining candidates are more difficult to classify. Therefore, a discriminative classifier SVM is used to classify the difficult cases which can not be classified in the first stage.

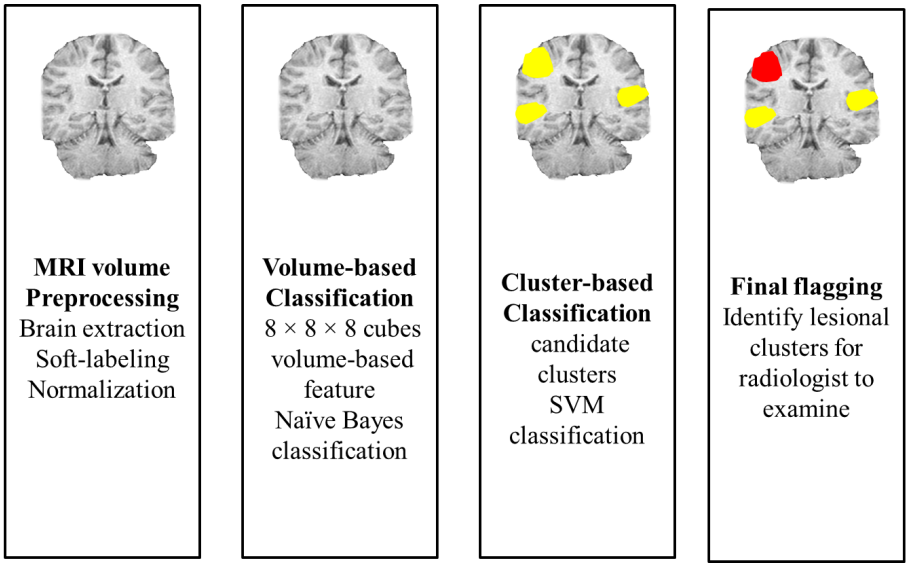


Figure 3.6: Proposed procedures for automated FCD analysis T1-weighted MRI

# Chapter 4

# Automated FCD Lesion Segmentation of T2 Weighted MRI

In this chapter, we introduced a new CAD system of FCD lesion detection on T2 weighted FLAIR MRI based on Markov Random Field with a location prior designed specifically for FCD lesions applications. In Section 4.1, we discuss the two pre-processing steps: bias field correction and mutual information registration. These pre-processing steps are essential for the subsequent analysis. In Section 4.2, the Markov Random Field framework used in this study is described. In Section 4.3, the energy minimization problem from Section 4.2 is formulated in the lesion labeling problem. The energy function and prior function are also defined in this section. In Section 4.5, we use graph cut to minimize the energy function, and in Section 4.6, the chapter is summarized.

## 4.1 Pre-processing

### 4.1.1 Bias field correction

A number of pre-processing procedures are needed for further analysis. The first step of pre-processing is the correction of intensity inhomogeneities of volumetric MRI scans. Given the intra-class intensity homogeneity assumption of most intensity-based tissue classifiers, the major difficulty in the segmentation of MR images is the intensity inhomogeneity artifact, and therefore the correction of this intensity variation is very crucial for achieving good segmentation results. Here we used the most widely used method for automatic correction of the bias field: Non-parametric Non-uniform intensity Normalization (N3), which is proposed by Sled et al [79]. The basic principles and methodologies of N3 are described in this section. First, let us consider the following model for MR image formation:

$$v(\mathbf{x}) = u(\mathbf{x})f(\mathbf{x}) + n(\mathbf{x}), \tag{4.1}$$

where at location $\mathbf{x}$, $v$ is the measured signal, $u$ is the true signal emitted by the tissue, $f$ is an unknown bias field, and $n$ is white Guassian noise. Consider $n(\mathbf{x}) = 0$ at each voxel location $\mathbf{x}$, using logarithmic notation $\hat{u}(\mathbf{x}) = \log(u((x)))$ the image formation model becomes additive:

$$\hat{v}(\mathbf{x}) = \hat{u}(\mathbf{x}) + \hat{f}(\mathbf{x}). \tag{4.2}$$

Let $U$, $V$, and $F$ be the probability densities of $\hat{u}$,$\hat{v}$, and $\hat{f}$ respectively. Assume that $\hat{u}$ and $\hat{f}$ are independent random variables, then the density of their sum is the convolution if the two,

$$V(\hat{v}) = F(\hat{v}) * U(\hat{v}) = \int F(\hat{v} - \hat{u})U(\hat{u})d\hat{u}. \tag{4.3}$$

Given the distribution $U$, the method of estimating the bias field is to treat measurement $\hat{v}$ as random samples. To estimate $\hat{u}$ based on the measurement $\hat{v}$ using expected value:

$$E[\hat{u}|\hat{v}] = \int_{-\infty}^{\infty} \hat{u}p(\hat{u}|\hat{v})d\hat{u}. \tag{4.4}$$

An estimate of $\hat{f}$ can be obtained using the estimate of $\hat{u}$ from eq(4.4) as follows,

$$\hat{f}_e(\hat{v}) = E[\hat{f}|\hat{v}] = \hat{v} - E[\hat{u}|\hat{v}], \tag{4.5}$$

where $\hat{f}_e$ is an estimate of $\hat{f}$ at location $\mathbf{x}$ based on the $\hat{v}$. The estimate can be smoothed by operator $S$. This estimation keeps iterating itself until it converges to within a certain error range that is predefined as our initial estimate of the field. An example of the bias field correction is presented in Fig 4.1.

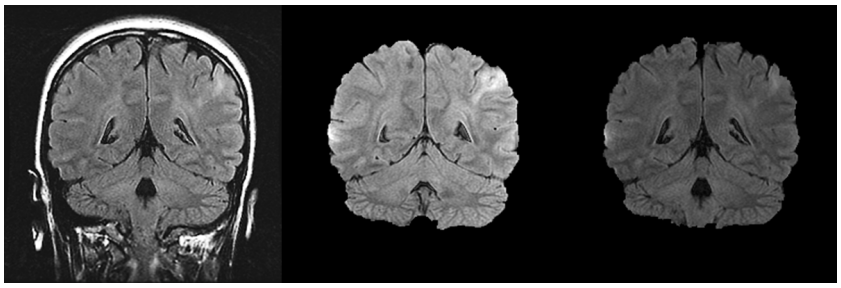

Figure 4.1: Example of using N3 Bias field corrections (left) original image (middle) brain extraction (right bias field corrected)

### 4.1.2 Mutual information Multi-modality Registration

Multi-modal medical image registration is another important pre-processing step for our segmentation task in order to combine the complementary information presented in T1 and T2 images. In this thesis, the well known Viola and Wells mutual information based for multi-modalities image registration method [80] is used. The derivation of the Viola-Wells mutual information registration is briefly presented in the following paragraphs.

Assume a voxel in the moving image data to be $v(x)$ and a voxel in the target image data to be $u(x)$, where $x$s are the voxel locations. Given a transformation $T$ that transform the coordinate frame $u(x)$ of the target volume to the test volume $v(x)$, then $v(T(x))$ is the same as $u(x)$. The mutual information based registration is seeking a transformation that maximizes the mutual information $I$ between the registered volume $u$ and the target volume $v$,

$$\hat{T} = arg\max_{T} I(u(x), v(T(x))), \tag{4.6}$$

where the definition of mutual information is based on the entropy theory which can be written as:

$$I(u(x), v(T(x))) \equiv h(u(x)) + h(v(T(x))) - h(u(x), v(T(x))), \tag{4.7}$$

where $h(\cdot)$ is the entropy of random variables, and is defined as,

$$h(x) \equiv -\int p(x) \ln p(x) dx, \tag{4.8}$$

while the joint entropy of two random variables $x$ and $y$ is

$$h(x, y) \equiv -\int p(x, y) \ln p(x, y) dx dy. \tag{4.9}$$

To be able to perform the maximization problem in Eq(4.6), several terms require estimation from the data. One needs to estimate the entropies and their derivatives, and also the derivatives of mutual information. These estimations and calculations are shown in the following.

*Density Estimation* To estimate entropy from a given image volume, one needs to first approximate the underlying probability density $p(z)$. By using *Parzen Window* density estimate [81], we have the following approximation:

$$p(z) \approx P^*(z) \equiv \frac{1}{N_A} \sum_{z_j \in A} R(z - z_j), \tag{4.10}$$

where $N_A$ is the number of trials in the sample $A$ and $R$ is a window function which integrates to 1. Similarly, the entropy integral can be approximated as,

$$h(z) \approx -\frac{1}{N_B} \sum_{z_i \in B} \ln P^*(z_i), \tag{4.11}$$

where $N_B$ is the size of the second sample $B$. We may now write an approximation for the entropy of random variable $z$ as,

$$h(z) \approx h^*(z) \equiv \frac{-1}{N_B} \sum_{z_i \in B} \ln \frac{1}{N_A} \sum_{z_j \in A} G_\psi(z_i - z_j), \tag{4.12}$$

where $G_\psi$ is a Gaussian density function. In order to find the maxima of entropy or mutual information, one needs the derivative of such functions with respect to the transformation $T$. After some manipulation, the derivative of the entropy can be written as follows,

$$\frac{d}{dT}h^*(v(T(x))) = \frac{1}{N_B}\sum_{x_i\in B}\sum_{x_j\in A} W_v(v_i, v_j)(v_i - v_j)^T\psi^{-1}\frac{d}{dT}(v_i - v_j), \qquad (4.13)$$

where $v_i$ is $v(T(x_i))$ and

$$W_v(v_i, v_j) \equiv \frac{G_{\psi_v}(v_i - v_j)}{\sum_{x_k\in A} G_{\psi_v}(v_i - v_k)}.$$

Therefore we can evaluate mutual information using the entropy estimation obtained in Eq(4.12). To seek maximum mutual information, one needs to calculate its derivative. After some manipulation, we have,

$$\widehat{\frac{d}{dT}I(T)} \approx \frac{d}{dT}h^*(u(x)) + \frac{d}{dT}h^*(v(T(x))) - \frac{d}{dT}h^*(u(x), v(T(x))). \qquad (4.14)$$

Now we have the approximated gradient information of the mutual information with respect to the transformation, which is the necessary information for gradient descent optimization method. In words, the mutual information registration seeks a local maximum of mutual information by using approximations of gradient descent. Each optimization step is repeatedly taken to update the respective transformation until convergence. The summary of the mutual information based registration is as follows:

| Repeat: | |
|---|---|
| $A \leftarrow$ | { sample of size $N_A$ drawn from $x$ } |
| $B \leftarrow$ | { sample of size $N_B$ drawn from $x$ } |
| $T \leftarrow$ | $T = T + \lambda\widehat{\frac{dI}{dT}}$ |

Where the parameter $\lambda$ is called the learning rate. The above procedures can be repeated until the stopping criterion is met. The stopping criterion can either

be a fixed number of times or a predefined definition of convergence is detected. An example of a pair of registered T1 and T2 FLAIR images using Viola-Wells mutual information based registration is presented in Fig 4.2.

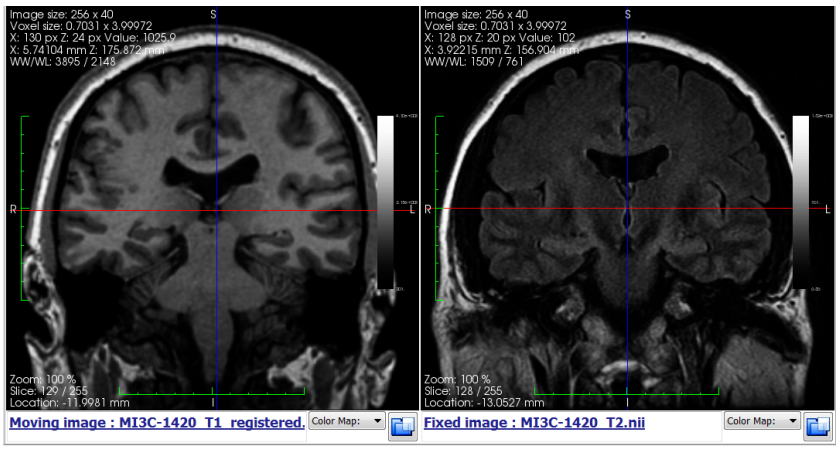


Figure 4.2: Viola-Wells multi-modality registration example result: registered T1 (left) and T2 (right) images

## 4.2 Markov Random Field Modeling

After the preprocessing steps described in the previous sections, including bias field correction, Viola-Wells mutual information based multi-modalities image registrations, the T1 and T2 weighted MRI images are now intensity-homogenous and spatially aligned for further analysis. In this section, we model the intensities of T2-weighted FLAIR MRI image using Markov Random Field Modeling.

Markov Random Field Modeling (MRF) has been used widely in the field of image understanding for its ability to model spatial coherence and stochastic interaction in the both the image and feature space. It has been proved to be quite useful in many practical applications, therefore tremendous attentions were generated especially in the literature of image processing, computer vision, and applied statics [82] using MRF. Recently, MRF modeling has also received much attention in the medical image analysis field, for its ability to embed various prior

information [83]. In the following sections, We briefly introduce the theory of MRF modeling [66] and how it can be directed into the application of the FCD lesion segmentation problem that we are interested.

### 4.2.1 Markov Random Field

Markov Random Field is an $n$-dimensional random process defined on a set of lattice points $S$. For every lattice point $s \in S$ with value $X_s$, there is a neighborhood system $\partial s$. A clique on a lattice is a set of points $c$, which are all neighbors of each other, and the definition of a clique can be shown as:

$$\forall s, r \in c, r \in \partial s. \tag{4.15}$$

The definition of Markov Random Field is for a random object $X$ on the lattice $S$ with neighborhood system $\partial s$,

$$p(x_s|x_r, for\ r \neq s) = p(x_s|x_{\partial r}), \tag{4.16}$$

which is true for all $s \in S$. Using Hammersley-Clifford Theorem [84] asserts that if $X$ is a Markov Random Field and $\forall x, P\{X = x\} > 0$ then $P\{X = x\}$ has the form of a Gibbs distribution. A probability distribution of $X$ is called a Gibbs distribution if it can be written as the following form,

$$p(x) = \frac{1}{Z} exp\{-\sum_{c \in C} V_c(x_c)\}. \tag{4.17}$$

Where $C$ is the set of all cliques and $Z$ is the normalizing constant for the density. $V_c(x_c)$, the potential function, is the any function of $x_c$. $U(x) = \sum_{c \in C} V_c(x_c)$ is known as the energy function. The energy function, also called the *prior energy* has the following form:

$$U(x) = \sum_{c \in C} V_c(x_c) = \sum_{\{i\} \in C_1} V_1(x_i) + \sum_{\{i,j\} \in C_2} V_2(x_i, x_j) + ..., \tag{4.18}$$

where '...' denotes higher order term. The potential functions contain *a priori* knowledge of how the values within the neighboring system interact with each

other. For any realization of $X = x$, we can derive the likelihood function for observation $\mathbf{r}$, which can be expressed as the exponential form,

$$p(\mathbf{r}|X = x) = \frac{1}{Z_r} \times exp\{-U(\mathbf{r}|x)\}, \tag{4.19}$$

where $U(\mathbf{r}|x)$ is called the *likelihood energy*. Then the posterior probability is also a Gibbs distribution,

$$P(X = x|\mathbf{r}) = \frac{1}{Z_E} \times exp\{-E(x)\}. \tag{4.20}$$

with *posterior energy* $E(x)$ which is

$$E(x) = U(x|\mathbf{r}) = U(x) + U(\mathbf{r}|x). \tag{4.21}$$

The Maximum a priori (MAP) solution for estimating parameter $X$ is equivalently found by

$$X^* = \arg\min_{X \in S} U(X|\mathbf{r}). \tag{4.22}$$

To summarize, the MRF modeling of the underlaying spatial distribution consists the following steps:

- Define a neighborhood system $S$, cliques $C$
- Define of prior information $U(x)$
- Derive likelihood energy $U(x|\mathbf{r})$ and the posterior energy $U(\mathbf{r}|x)$

.

In this section, we introduce MRF modeling and how it can govern the underlying distributions using a spatial neighborhood system (clique) and prior information, which can greatly simplify the estimation problem with the help of Gibbs distribution. In the next section, we cast the FCD lesion segmentation task in T2 weighted MRI as a solution of label configuration under the assumption of MRF.

## 4.3 Markov Random Field Segmentation

### 4.3.1 Lesion segmentation problem as labeling

Our task of FCD lesion segmentation on T2 weighted MRI images can be seen as a labeling problem. A *labeling problem* can be described as a set of sites (e.g. image pixels, voxels) and a set of labels(e.g. healthy, lesion, or tissue types). Let us assume $\mathbf{d}$ be a set of $m$ sites

$$\mathbf{d} = \{d_1, d_2, d_3, ..., d_m\}. \tag{4.23}$$

Also, let $\mathbf{L}$ be a set of labels. Labeling is to assign a label from $\mathbf{L}$ to each of the sites in $\mathbf{d}$. In our case, $\mathbf{d}$ is the set of voxels and $\mathbf{L}$ is the set of class labels, which in this study, is either lesional or healthy. Let $X = \{X_1, X_2, ...X_m\}$ be random variables defined on each voxel in $\mathbf{d}$ which each variable $x_i$ assumes a value in $\mathbf{L}$, represents the intrinsic class label. A realization of $X$ where $x = \{x_1, x_2, ...., x_m\}$ is called a configuration of $X$. A configuration here represent a mapping from the image features to the class label. From the previous section, we know that the best labeling or realization of configurations is achieved by minimizing the posterior energy $U(X|\mathbf{r})$ by given the observation $\mathbf{r}$, as in eq.(4.22). Therefore, our next step is to characterize and define this posterior energy for the segmentation by incorporate both the neighborhood system and prior information. However, before we define our prior information, we first introduce the Expectation Maximization (EM) algorithm that is essential for constructing our prior information.

## 4.4 Definition of the segmentation energy $U(x|\mathbf{r})$

### 4.4.1 Model Fitting Using EM algorithm

A Gaussian Mixture Model (GMM) or Finite Normal Mixture (FNM) [85] is often assumed that for the pixel intensity $r_i$ of each site $d_i$. The observed random

variables have the following density function:

$$p(r|x) = \sum_{l \in \mathbf{L}} w_l \cdot g(r; \theta_l) \tag{4.24}$$

where $\theta_l = (\mu_l, \sigma_l)^T$ and

$$g(r; \theta_l) = \frac{1}{\sqrt{2\pi\sigma_l^2}} exp(-\frac{(y-\mu_l)^2}{2\sigma_l^2}). \tag{4.25}$$

In practice, the parameter set $\theta = \theta_l, l \in L$ and the class label are both unknown and needs to be determined. A lot of work has been proposed to solve this problem, among all of them, Expectation-Maximization (EM) [86] algorithm is the one that is the most widely used. The techniques and strategies underlaying the EM algorithm is described in the following.

First we start with an initial estimate of the set parameters $\theta^{(0)}$.

The *Expectation Step* or *E step* calculates the expected value of the log likelihood probability, which is a conditional distribution of $\mathbf{x}$ given $\mathbf{r}$ under the current estimate of the parameters $\theta^{(t)}$

$$Q(\theta|\theta^{(t)}) = E[\log P(\mathbf{x}, \mathbf{r}|\theta)|\mathbf{y}, \theta^{(t)}] = \sum_{x \in \mathbf{L}} p((x)|\mathbf{r}, \theta^t). \log p((x), \mathbf{r}|\theta) \tag{4.26}$$

The *Maximization Step* or *M step* maximizes $Q(\theta|\theta^{(t)})$ to obtain the next estimate

$$\theta^{(t+1)} = \arg\max_{\theta} Q(\theta|\theta^{(t)}). \tag{4.27}$$

Then $\theta^{(t+1)} \rightarrow \theta^{(t)}$ and repeat from the E step. The algorithm iterates until convergence. In the next section, the EM algorithm is used to provide estimate of one of the prior information.

### 4.4.2 Definition of The Prior Models

Here we define the segmentation energy or the posterior energy $U(X|\mathbf{r})$ as a combination of two energy functions, each of this function model a type of prior

information:

$$U(X|\mathbf{r}) = U^A(X|\mathbf{r}) + U^L(X|\mathbf{r}), \tag{4.28}$$

where the energy functions $U^A(X|\mathbf{r})$ model prior information about the appearance and $U^L(X|\mathbf{r})$ model the prior information about the locations. Each of these energy functions can be expressed as clique energies as defined in eq.(4.29), and we defined these functions in further details as follows,

1. Appearance prior $U^A(X|\mathbf{r})$ This energy function models the appearance or intensity value of the lesional and healthy area. The definition of $U^A(X|\mathbf{r})$ can be defined by using the $1^{st}$ and $2^{nd}$- order clique energies:

$$U^A(X|\mathbf{r}) = \sum_{i\in\mathbf{d}} V_i^A(x_i|\mathbf{r}) + \sum_{i\in\mathbf{d}}\sum_{j\in N_i} V_{ij}^A(x_i, x_j|\mathbf{r}), \tag{4.29}$$

   where $N_i$ is the neighboring pixels defined by the clique system on $\mathbf{d}$. $V_i^A(x_i|\mathbf{r})$ and $V_{ij}^A(x_i, x_j|\mathbf{r})$ are the $1^{st}$ and $2^{nd}$- order clique modeling the appearance information. In our study, the appearance information is modeled by the intensity value of the site (voxel). The $1^{st}$-order cinque potential $V_i^A(x_i|\mathbf{r})$ is defined as:

$$V_i^A(x_i|\mathbf{r}) = -\log[p(d_i = r_i; \theta_{x_i})], \tag{4.30}$$

   where $p(d_i = r_i; \theta_{x_i})$ is the probability distribution of the intensity value of the given class label $x_i$ (healthy or lesional), $\theta_{x_i}$ is a vector of parameters associated with the probability function and $r_i$ is the observed intensity value at the given voxel site $d_i$. For most of the work in medical image segmentation literature [83][85], a Gaussian Mixture Model (GMM)is often used. The parameters of the GMM can be learned using EM algorithm which is described in the pervious section.

   To introduction of the energy associated with the $2^{nd}$ order cliques help incorporate spatial information into the intensity based modeling. The most simple model to govern $2^{nd}$ order relationship is the Generalized Potts model [87] which can be shown as:

$$V_{ij}^{A}(x_i, x_j|\mathbf{r}) = B(i,j|\mathbf{r}).(1 - \delta_{x_i,x_j}), \tag{4.31}$$

where the function $\delta_{x_i,x_j}$ is the Kronecker delta function. $\delta_{x_i,x_j} = 1$ if $x_i = x_j$ and 0 otherwise. $B(i,j|\mathbf{r})$ is the penalty when the two adjacent sites do not agree with each other. The penalty $B(i,j|\mathbf{r})$ should also relate to the distance between the two sites $d_i$ and $d_j$ since natural images tend to have homogeneity among neighboring pixels. For general image segmentation application, we want the boundary to lie on the edges in the image, a typical choice for $B(i,j|\mathbf{r})$ is,

$$B(i,j|\mathbf{r}) = \beta \cdot exp(-\frac{(I_{d_i} - I_{d_j})^2}{2\delta^2} \cdot \frac{1}{dist(d_i, d_j)}), \tag{4.32}$$

where $I_{d_i}$ and $I_{d_j}$ are the intensity values of site $d_i$ and $d_j$ and $dis(d_i, d_j)$ is the distance between sites $d_i$ and $d_j$. Parameter $\delta$ in eq.(4.32) is related to the variation of the neighboring sites within the same class. When the constant $\beta = 0$, the function term $B$ is simply the traditional Ising prior. The constant $\beta$ chosen by Boykov et al [88] is:

$$\beta = (2\langle (I_{d_i} - I_{d_j})^2 \rangle), \tag{4.33}$$

where the $\langle \cdot \rangle$ denotes the expectation over an image sample.

Besides hyper-intensities on T2 FLAIR MRI, FCD lesions are not located at any part of the brain, on the other hand, it usually resides on the boundary of GM and WM. The *location prior* which we introduce next, help incorporate this information into our segmentation decision.

2. Location prior $U^L(X|\mathbf{r})$ The second energy function $U^L(X|\mathbf{r})$ in eq.(4.28) models prior information about the location. The information can be represented by using $1^{st}$ order clique potential as:

$$U^L(X|\mathbf{r}) = \sum_{i \in \mathbf{d}} V_i^L(x_i|\mathbf{r}), \tag{4.34}$$

where $V_i^L(x_i|\mathbf{r})$ is the penalty of assigning a given label $x_i$ to a given voxel $i$ according to the prior information about the location. This can be defined in a hard manner if some deterministic location information is provided (e.g. $V_i^L(x_i|\mathbf{r}) = \infty$ if $x_i$ is not in a previously labeled location). However, in this study, this approach is not suitable since we only know the lesions mostly locate at the GM and WM boundaries. Therefore, we define the location prior in soft-constrains manner, the location information is provided by T1 weighted MRI since it gives better soft tissue contrast. By doing the registration of the T2 and T1 weighted MRI, together with the thickness measurements [74] calculated in Section 3.2, we can construct the location prior function.

Instead of thickness, we use the distance of a given voxel $v$ to the outer surface which can be denote as $t_1(v)$, and the distance to inner surface as $t_2(v)$. (See Fig 4.3) The reason for selecting $t_1(v)$ instead of $t(v)$ is that we are interested in how the voxel is located related to the outer surface. Since FCD lesions tend to locate at the GM/WM boundaries, the greater $t_1(v)$ is, the more likely the voxel is lesional.

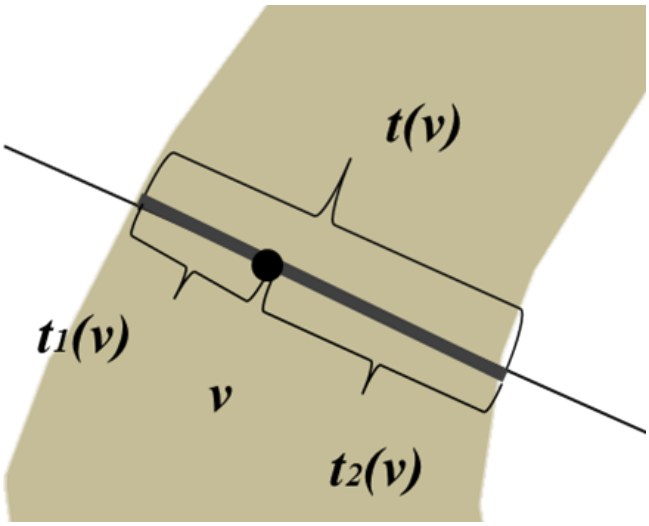


Figure 4.3: The thickness measurement at a given voxel $v$ is $t(v) = t_1(v) + t_2(v)$

In this study, we define $V_i^L(x_i|\mathbf{r})$ as follows: $V_i^L(x_i|\mathbf{r}) = w \cdot t_1(i)/t(i)$ if $x_i$ is healthy and $V_i^L(x_i|\mathbf{r}) = w \cdot (1 - t_1(i)/t(i))$ if $x_i$ is lesional. The term $t_1(i)$ is the distance from the outer surface of the cortex the location $i$, and is normalized with the thickness $t(i)$. The parameter $w$ can be defined by the user to control the importance of the location prior. The designed of $V_i^L(x_i|\mathbf{r})$ is to ensure that for lesional labeling, the greater $t_1(i)$, the lower the energy, meaning for voxels locate deeper in the cortex with hyper-intensity, the lesional labeling are encouraged.

In this section, we formulate the segmentation energy function using combination of appearance prior information and location prior information. In contrast to traditional methods, which only use appearance or intensity information, the introduction of location prior in this thesis provides additional information to assist segmentation. The segmentation problem is now an optimization problem of the energy function. In the next section, we discuss how the optimization problem is solved using Graph Cuts.

## 4.5 Graph Cuts

From the previous sections, we formulated the T2 FCD lesion segmentation or labeling problem into a minimization problem of an energy function $E$ in Eq.(4.21) so that the minimization corresponds to a good segmentation. Originally, MAP-MRF problem is generally solved by using simulated annealing (SA) or Iterated Conditional Mode (ICM). SA guarantees global minimal for any arbitrary energy function, however, it is very computationally expensive. On the other hand, ICM does not provide a globally best estimation. In 2001, Boykov [88] et. al proposed using graph cuts or max-flow/min-cut algorithm to solve the MAP-MRF and binary segmentation problem. In this section, we briefly describe the mapping of MAP-MRF to graph cut and how help solve the minimization problem more efficiently.

A undirected graph $G = \langle V, E \rangle$ can be defined by the composition of a set of vertices ($V$) and a set of undirected links called edges ($E$). Each edge $e \in E$ in the graph is given a cost called $w_e$. There are two terminal nodes called the *sink* and the *source*. A cut is a subgroup of edges $C \subset E$ such that the sink and source become separated if these edges are removed, and the induced graph can be written as $G(C) = \langle V, E \rangle$. The cost of this "cut" is the sum of all the costs $W_e$ of the edges,

$$|C| = \sum_{e \in C} w_e. \tag{4.35}$$

The graph cut formulation can be easily transformed into a binary image segmentation problem. The nodes can represent pixels or voxels and the edges can represent spatial neighborhood. A cut corresponds naturally to segmentation of the underlaying image or volume. We define $P$ as the set containing all the voxels $p$ of the brain and $N$ as the set containing all the connection between two nodes $\{p, q\}$, and segmentation is represented by $X$. Therefore eq. (4.28), the energy function can be re-written in the following format:

$$E(X) = \alpha \sum_{p} R_p(X_p) + \sum_{p,q \in N} B_{\{p,q\}}. \tag{4.36}$$

The term $R_p$ referred to the first order clique potential. In the graph, the relation is expressed by connection of all the nodes to the $t - links$: source and sink nodes, with weights $W^p_{source}$ and $W^p_{sink}$ respectively.

The boundary term $B_{\{p,q\}}$ referred to the similarity of the voxels $p$ and $q$, which is the $2^{nd}$ order clique potential in eq.(4.29). The coefficient $\alpha$ is used to balance the regional and boundary terms. This graph formulations enable us the use the graph minimization method proposed by Boykov et. al [88] which is designed to efficiently solve the energy minimization problem.

## 4.6 Summary

In this chapter, a CAD system to segment FCD lesions on T2 FLAIR images is proposed. After the preprocessing steps, which include bias field correction and mutual information registration of the T1 and T2 images, a Gaussian Mixture Model is used to model the intensity value of the T2 FLAIR images, estimated by the Expectation Maximization algorithm. This model is incorporated into a MRF modeling as the appearance prior. Besides the traditionally used appearance prior, we introduced a location prior based on the thickness measurements on T1 weighted MRI, which is known to provide a better contrast of the soft tissues for brain imaging. By combining both T2 hyper-intensity information and T1 location information, we integrate the advantages from both modalities to enhance the robustness of the segmentation.

# Chapter 5

# Statistical Analysis of Proposed Features

In this chapter, different statistical methods, including Quantile-Quantile (Q-Q) plot (Section 5.2.1), student's t-test (Section 5.2.2), and Receiver Operating Characteristic (ROC) analysis (Section 5.2.3), are used to evaluate the proposed 7 volume-based features. The statistical analysis is based on a clinical dataset with a total of 51 subjects with FCD lesions. The goal of this chapter is to demonstrate the usefulness and compare the performance of each of the proposed features in terms of diagnosising FCD lesions.

## 5.1 Data Description

A total of 51 subjects with diagnosis of FCD is included in this study. Scanning was carried out on General Electric (GE) Medical systems (GE Discovery, Genesis, Signa) of 1.5T (20 subjects) and 3T (31 subjects) at the Radiology Department, Mayo Clinic, Rochester. The image dimension is $256 \times 256 \times 256$ and voxel size is typically $0.86 \times 0.86 \times 0.77$ mm for T1 weighted scans. For T2 weighted images, the image dimension is $256 \times 256 \times 44$ and voxel size is $0.70 \times 0.70 \times 4$ mm. The subjects aged from 2 to 73 years. (Mean= 30.23 years, SD= 15.58 years, 23 males,

28 females) The summary of the information of subjects is in Table 5.1.

## 5.2 Volume-based Feature Statistics Analysis

The seven volume-based features designed to characterize FCD lesion on T1 weighted MRI, were calculated for each $8 \times 8 \times 8$ cube from the brain volume of each subject. The cube is labeled as "healthy" if less than 10% of its total voxels (less than 51 lesional voxels in a cube for our study) are lesional. Lesional areas were identified by at least two radiologist and labeled by a trained personnel; otherwise the cube is labeled as "lesional". The calculated numerical averages are summarized in Table 5.2. From the table, we observe several anticipated facts. First, the average cortical thickness is larger at lesional areas. Second, the absolute gradient is smaller in lesional cubes. These are the known characteristics of FCD, which are successfully captured using the proposed features. Also, the variance of the gradient vectors, which we assumed to represent randomness and blurriness, are larger for the lesions. For thickness skewness, the healthy thickness distribution has a positive skew, indicating a distribution concentrated on the smaller end; the lesional distribution has a negative skew, suggesting that values lie to the right of the mean. This also complies with our initial assumption. For both kurtosis measures for thickness and gradient vectors, the healthy cubes preserve a higher kurtosis, which can translate as high peakedness and thinner tails, which the distributions of lesional cubes represent more flatness.

### 5.2.1 Quantile-Quantile (Q-Q) plot

In statistics, a Q-Q plot [89] is a probability plot, which is a graphical method for comparing two probability distributions by plotting their quantiles against each other. Q-Q plot can assist our study of the distributions of the proposed features by visual illustrations., which can provide a graphical view of the shape of two

Table 5.1: Subject Demographic

| Subject | Gender (M,F) | Age(years) | Lesion Size($mm^3$) | Magnetic Field | Lesion location | T1 | T2 |
|---|---|---|---|---|---|---|---|
| 1 | M | 2 | 1287 | 1.5T | left frontal | + | + |
| 2 | M | 3 | 577 | 1.5T | left frontal | + | + |
| 3 | F | 4 | 450 | 1.5T | Inf. L Occip | + | + |
| 4 | M | 8 | 3688 | 1.5T | right frontal | + | + |
| 5 | M | 9 | 5239 | 1.5T | R parietal and frontal | + | + |
| 6 | M | 10 | 375 | 3T | R occip | + | + |
| 7 | M | 13 | 434 | 3T | L mid frontal lobe | + | + |
| 8 | F | 16 | 1030 | 1.5T | L mid frontal gyrus | + | + |
| 9 | M | 14 | 2455 | 3T | R frontal operculum | + | + |
| 10 | F | 24 | 3429 | 1.5T | right frontal | + | + |
| 11 | M | 26 | 1621 | 3T | R inf temp gyrus | + | + |
| 12 | M | 23 | 2004 | 1.5T | Bilat sup frontal gyri | + | + |
| 13 | M | 24 | 1234 | 1.5T | Post L sup frontal gyrus | + | + |
| 14 | F | 24 | 2896 | 1.5T | Left insula | + | + |
| 15 | F | 25 | 1834 | 3T | R sup frontal gyrus | + | + |
| 16 | M | 27 | 2610 | 3T | R parietal lobe | + | + |
| 17 | M | 37 | 118 | 1.5T | L frontal lobe | + | + |
| 18 | F | 40 | 4230 | 1.5T | R temp lobe superomesial | + | + |
| 19 | M | 16 | 7245 | 1.5T | left frontal centrum semiovale | + | + |
| 20 | M | 31 | 1921 | 3T | right frontal | + | + |
| 21 | F | 46 | 987 | 1.5T | L frontal operculum | + | + |
| 22 | F | 58 | 2407 | 3T | R temp parietal | + | + |
| 23 | F | 54 | 230 | 1.5T | inf and lat L temp lobe | + | + |
| 24 | M | 65 | 982 | 3T | Post-inf L temp lobe | + | + |
| 25 | M | 73 | 2396 | 1.5T | Post R frontal lobe | + | + |
| 26 | M | 44 | 1002 | 1.5T | left frontal lobe | + | + |
| 27 | F | 45 | 5317 | 1.5T | R post parietal | + | + |
| 28 | M | 49 | 532 | 3T | R frontal-parietal | + | + |
| 29 | M | 9 | 911 | 3T | R temp par | + | + |
| 30 | F | 31 | 2063 | 3T | left frontal | + | + |
| 31 | F | 20 | 1216 | 3T | L temp and insula | + | + |
| 32 | F | 27 | 209 | 3T | R sup. and mid temp | + | + |
| 33 | F | 40 | 410 | 3T | R sup temp | + | + |
| 34 | M | 43 | 6456 | 3T | R ant temp and amydgala | + | + |
| 35 | F | 28 | 1442 | 3T | L temp and L occ | + | + |
| 36 | M | 43 | 1777 | 3T | left temporal lobe | + | + |
| 37 | M | 9 | 2817 | 3T | L inf temp | + | + |
| 38 | M | 46 | 2376 | 3T | R temp | + | + |
| 39 | F | 48 | 2817 | 3T | R inf frontal | + | + |
| 40 | M | 27 | 54 | 3T | R sup frontal | + | + |
| 41 | M | 48 | 4903 | 3T | R sup temp | + | + |
| 42 | F | 25 | 5155 | 3T | left frontal lobe | + | + |
| 43 | M | 23 | 641 | 3T | Post-inf Lt temp lobe | + | + |
| 44 | F | 35 | 1382 | 3T | right frontal gyrus | + | + |
| 45 | F | 33 | 114 | 3T | Sup L frontal lobe | + | + |
| 46 | M | 26 | 5553 | 3T | R post parietal | + | + |
| 47 | M | 26 | 2014 | 3T | post R temp lobe | + | + |
| 48 | M | 18 | 1139 | 3T | Bilat frontal | + | + |
| 49 | F | 31 | 11011 | 3T | Rt temp parietal | + | + |
| 50 | M | 20 | 218 | 3T | left frontal | + | + |
| 51 | M | 28 | 1258 | 3T | left frontal | + | + |

Table 5.2: Average feature values for healthy and FCD lesional volumetric cubes

| Feature | Lesional | Healthy |
|---|---|---|
| Cortical Thickness (cm) | 4.1909 | 3.4574 |
| Thickness Skewness | -0.2125 | 1.0371 |
| Thickness Kurtosis | 1.3232 | 4.2254 |
| Absolute Gradient | $4.90 \times 10^3$ | $5.40 \times 10^3$ |
| Gradient Vectors Variance | 2.7608 | 2.4082 |
| Gradient Vectors Skewness | -0.4044 | 0.077 |
| Gradient Vectors Kurtosis | 2.0048 | 2.6308 |

different distributions Here we represent the histograms and Q-Q plots of subject #1 as a typical example. In Fig 5.1, the histograms of the three fundamental features are plotted for both healthy and lesional labeled cubes. It is very clear that for thickness distribution, the histogram for healthy regions has a skew to the left (positive skew) and, on the other hand, for a lesional area, the histogram has a skew to the right(negative skew). The skewness is displayed similarly for the feature of variance of gradient vectors, where the lesional cubes represent a negative skew compared to the healthy ones. On the contrary, one can see from the middle row of Fig 5.1 that for the gradient distributions, although the mean values of the two distributions are clearly different, the shape of the two distributions both resemble a normal distributions, which no higher tails on either side. The observations we made above can also be demonstrated using Q-Q plots. In Fig 5.2, the first row represents the Q-Q plot of the thickness measure distributions. From left to right, the first plot represents thickness distributions of healthy tissue (Thickness(H)) compared with a Normal distribution, and the second one is lesional distribution (Thickness(L)) compared with Normal distributions. The third plot is the Q-Q plot of thickness distributions of healthy and lesional areas. From these Q-Q plots, it is obvious that the two distributions are different, which is especially clear in the third figure when we compare the two distributions directly. One can see that the curved pattern suggests that the thickness distribution from the lesional area are more skewed to the right. Similar curved patterns can be observed in the Q-Q plots of the variances of gradient for lesional (Gradient(L))

and healthy (Gradient(H)) areas. For the Q-Q plots of gradient distributions, we can observe that both distributions of Gradient(H) and Gradient(L) agree with the Normal distribution and each other.

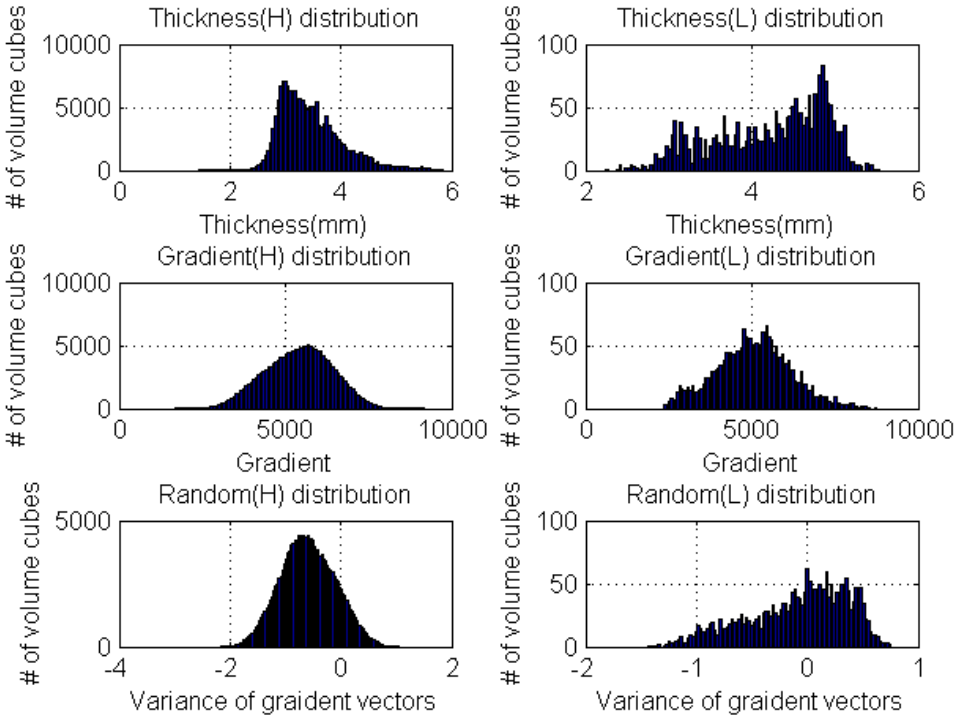


Figure 5.1: The histograms of subject #1: (top)thickness, (middle)gradient, and (bottom)variance of gradient features distributions

### 5.2.2 Student's t-test

A *t-test* or a *student's t test* [90] is any statistical hypothesis test in which the test statistics follows a student's t distribution if the null hypothesis is true. For testing hypothesis using the student's t-test, the sampled distribution is assumed to be Gaussian. However, as shown from the previous Section (5.2.1), our sampled

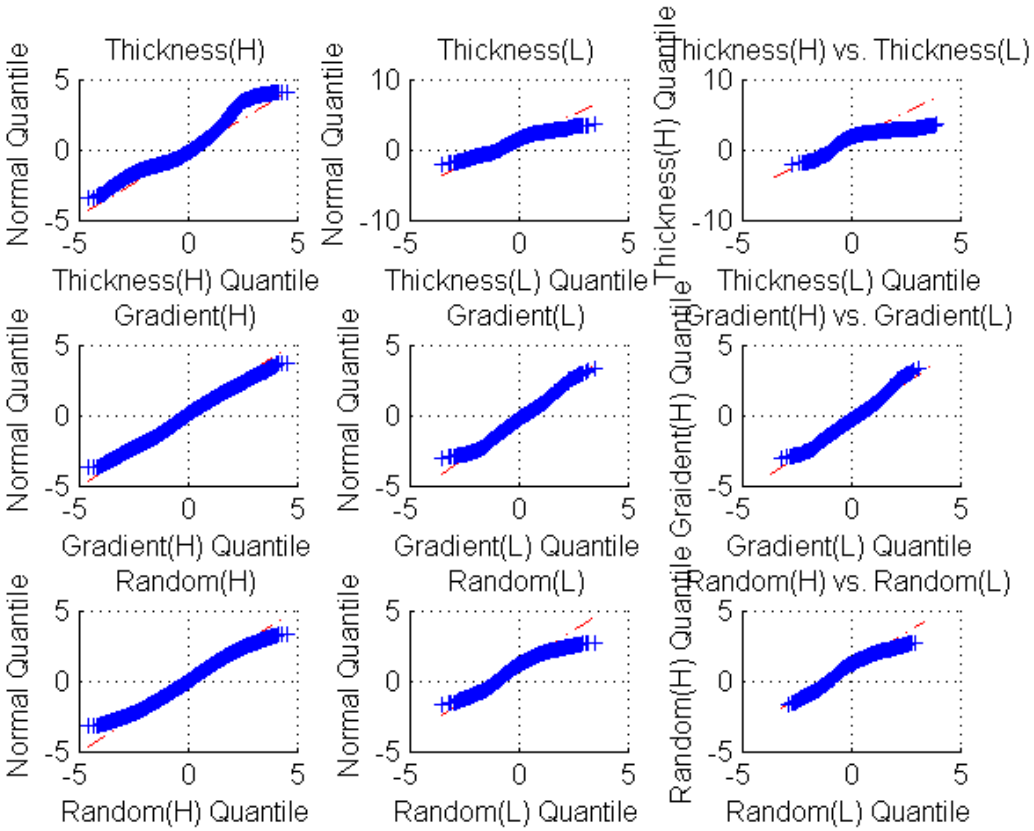


Figure 5.2: The Q-Q plot of subject #1 thickness, gradient, and variance of gradient feature distributions

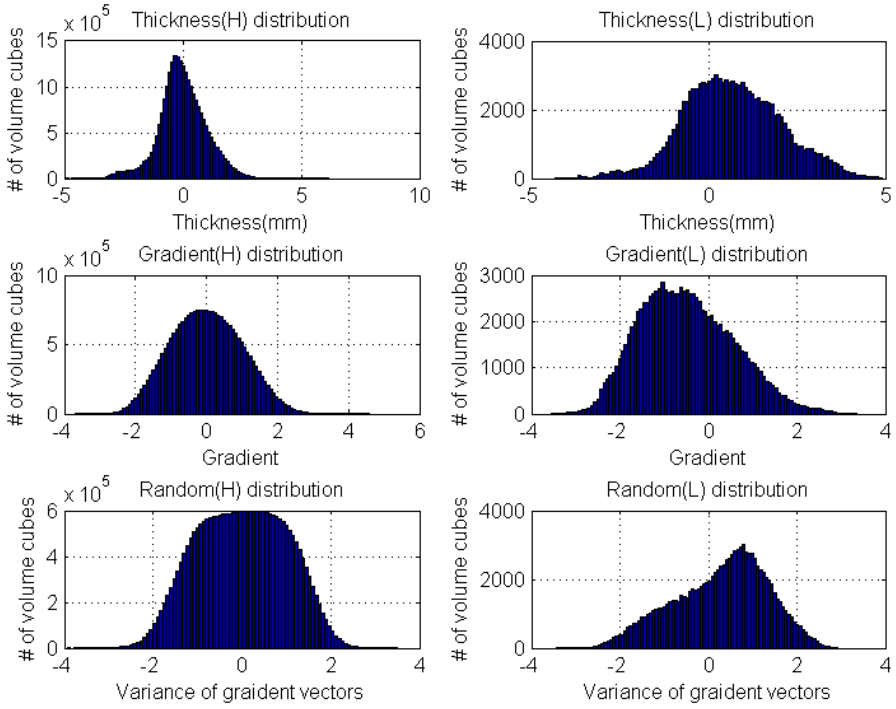


Figure 5.3: The histograms of all subjects: (top)thickness, (middle)gradient, and (bottom)variance of gradient features distributions

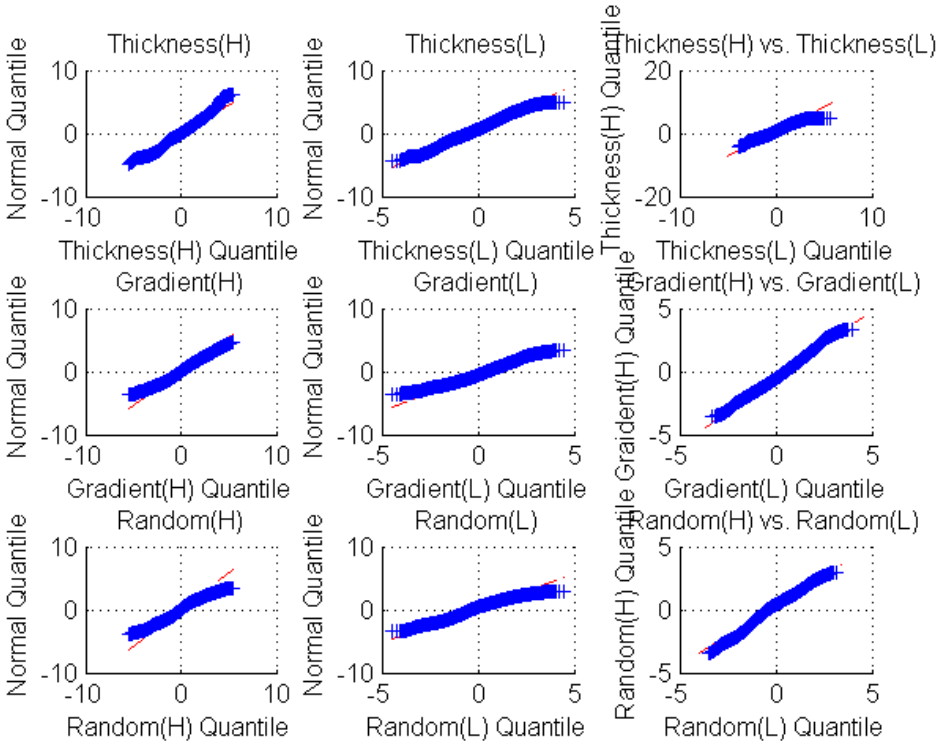


Figure 5.4: The Q-Q plot of all subjects for thickness, gradient, and variance of gradient feature distributions

distribution is not Gaussian. In this situation, one can use Central Limit Theorem, because the number of samples (1000-5000 samples for lesional distributions, and hundreds of thousands for healthy distribution) is very large, the mean of any distribution is approximately Gaussian.

In our experiment, we use a un-paired t-test with unequal sample sizes and unequal variances for two samples $X_1$ and $X_2$, which is also known as the *Welch's t test* The t-statistics can be calculated as follows,

$$t = \frac{\overline{X}_1 - \overline{X}_2}{S_{\overline{X}_1 - \overline{X}_2}}, \tag{5.1}$$

where,

$$S_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \tag{5.2}$$

with$s^2$ as unbiased estimator of the variance and $n$ the number of samples from the group. For the use of significance testing, the degree of freedom is calculated using

$$d.f. = \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{(\frac{s_1^2}{n_1 - 1})^2 + (\frac{s_2^2}{n_2 - 1})^2}. \tag{5.3}$$

In statistical significance testing, the *p-value* is often used for the evaluation of significance. The definition of p-value is he probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. Traditionally, one rejects the null hypothesis if the p-value is less than or equal to the significance level which is often represented by the Greek letter $\alpha$. To test if the features help differentiate lesional and healthy area, we apply t-test on the samples of each feature distributions from both lesional and healthy area and comparing the p-value. In other words, t-test can be used as a measure of how effective the feature is at separating groups. In Table 5.3, we summarize the t-test and the $p$-value for each of the features. From Table 5.3, all of the features have $p$-values equal zero, or a very small p-value that is lower than the number precision of the computer. Therefore, from the statistics of the entire dataset, the features indeed have the effectiveness of for FCD lesion

detection. However, we are not only interested in the effectiveness of features from the overall data, but more importantly, the effectiveness of features in an individual subject. We present subject #50 as an example of some $p$-value above the typical significant level $\alpha = 0.05$ in Table 5.4. As we can see from Table 5.4, 2 out of 7 features fall below $\alpha$: the kurtosis of the thickness and absolute gradient vectors. To demonstrate the effectiveness of the features, we apply the t-test for each individual subject, record the incident whenever there is a feature's $p$-value falls below $\alpha$, and the results are summarized in Table 5.5. From Table 5.5, all features have effectiveness around or above 90% of the subjects. Among all features, the thickness feature only fails in one subject, and the second best is our proposed gradient vector variance, which performs more consistently than the traditionally used absolute gradient. Table 5.5 clearly shows that despite the variability and subtlety in our dataset, the feature distributions should be statistically different between healthy and lesional groups. Note that for the t-test fails to be significant, it may also imply that certain type of lesion characteristics are not fully represented in the subjects. Studies have shown that the three main characteristics of FCD do not necessarily always coexist [43].

Table 5.3: T-test $p$-values for healthy and FCD lesional volumetric cubes

| Feature | Lesional | Healthy | t-test p-value |
|---|---|---|---|
| Cortical Thickness (cm) | 4.1909 | 3.4574 | 0 |
| Thickness Skewness | -0.2125 | 1.0371 | 0 |
| Thickness Kurtosis | 1.3232 | 4.2254 | 0 |
| Absolute Gradient | $4.90 \times 10^3$ | $5.40 \times 10^3$ | 0 |
| Gradient Vectors Variance | 2.7608 | 2.4082 | 0 |
| Gradient Vectors Skewness | -0.4044 | 0.077 | 0 |
| Gradient Vectors Kurtosis | 2.0048 | 2.6308 | 0 |

Table 5.4: T-test $p$-values for healthy and FCD lesional of subject #50

| Feature | Lesional | Healthy | t-test p-value |
|---|---|---|---|
| Cortical Thickness (cm) | 4.9909 | 1.14574 | 0.0154 |
| Thickness Skewness | 0.1705 | 0.6459 | 0.0453 |
| Thickness Kurtosis | 3.0532 | 6.0524 | 0.0595 |
| Absolute Gradient | $5.90 \times 10^3$ | $2.82 \times 10^3$ | 0.0361 |
| Gradient Vectors Variance | 1.028 | 2.229 | 0.0295 |
| Gradient Vectors Skewness | -0.2232 | 0.8202 | 0.0361 |
| Gradient Vectors Kurtosis | 2.2240 | 2.5758 | 0.1625 |

Table 5.5: Feature effectiveness by t-test $p$-value with $\alpha = 0.05$

| Feature | # subjects above $\alpha$(%) | # subjects below $\alpha$(%) | average $p < 0.05$ |
|---|---|---|---|
| Cortical Thickness (cm) | 1 (1.96%) | 50(98.04%) | $0.7656 \pm 0$ |
| Thickness Skewness | 3 (5.88%) | 48(94.12%) | $0.3236 \pm 0.2007$ |
| Thickness Kurtosis | 5 (9.80%) | 46(90.20%) | $0.2390 \pm 0.1631$ |
| Absolute Gradient | 4 (7.84%) | 47(92.16%) | $0.5167 \pm 0.1901$ |
| Gradient Vectors Variance | 3 (5.88%) | 48(94.12%) | $0.2130 \pm 0.1631$ |
| Gradient Vectors Skewness | 4 (7.84%) | 47(92.16%) | $0.4521 \pm 0.3098$ |
| Gradient Vectors Kurtosis | 6 (11.76%) | 45(88.24%) | $0.3419 \pm 0.3427$ |

### 5.2.3 Receiver Operating Characteristic (ROC) Analysis

Receiver Operating Characteristic (ROC) curves are plots of sensitivity, or true positive rates versus false positive rates. ROC was originally applied in signal detection theory [91], and has been extended its applications to several different fields include medicine, radiology and biometrics, which it can serve as a evaluations of accuracy versus risk. Some statistics, such as Area Under Curve (AUC), which associated with ROC curves can also be used in the realm of feature ranking and selections [92]. The AUC is usually calculated using normalized units, which can be intuitively transform as the probability will label a randomly chosen positive instance higher than a randomly chosen negative one. The maximum value for AUC is 1.0, thereby indicating a perfect test. An AUC value 0.5 indicates no discriminative value, as its performance is equal to a random quess. Consider a binary classification task with $m$ positive examples and $n$ negative examples. Assume that a classifier outputs a strictly ordered list for these examples and will denote by $1_X$ the indicator function of a set $X$. Let $c$ be a fixed classifier. Let $x_1, ..., x_m$ be the output of $c$ on the positive examples and $y_1, ..., y_n$ its output on the negative examples. Then, the AUC associated with $c$ is given by:

$$A = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} 1_{[x_i > y_i]}}{mn}. \tag{5.4}$$

Besides AUC, a typical binary classification test can also be measured using *sensitivity* and *specificity*. *Sensitivity* measures the proportion of actual positives which are correctly identified as such. *Specificity* measures the proportion of negatives which are correctly identified. The definitions are given as follows,

$$sensitivity = \frac{number\ of\ True\ Positives}{number\ of\ True\ Positives\ +\ number\ of\ False\ Negatives}. \tag{5.5}$$

and

$$specificity = \frac{number\ of\ True\ Negatives}{number\ of\ True\ Negatives\ +\ number\ of\ False\ Positives}. \tag{5.6}$$

To demonstrate the usefulness of a feature with the ROC curve and AUC, we randomly select 5000 feature vectors from subject #50 and summarized the results in Table 5.6 and their corresponding ROC curves can be seen in Fig 5.5 . The sensitivity and specificity are selected at the optimal point (The one closest to [0,1]). The traditional guide of classifying the accuracy of a diagnostic test as follows:

- *AUC Above 0.9* - excellent test
- *AUC = 0.8-0.9* - good test
- *AUC = 0.7-0.8* - fair test
- *AUC =0.6-0.7* - poor test
- *AUC =0.5-0.6* - fail test

Table 5.6: ROC statistics for proposed features in subject #50

| Feature | AUC | sensitivity | specificity |
|---|---|---|---|
| Cortical Thickness (cm) | 0.7694 | 0.7652 | 0.7673 |
| Thickness Skewness | 0.8592 | 0.7652 | 0.749 |
| Thickness Kurtosis | 0.9504 | 0.9217 | 0.8684 |
| Absolute Gradient | 0.8753 | 0.8522 | 0.818 |
| Gradient Vectors Variance | 0.9315 | 0.9217 | 0.831 |
| Gradient Vectors Skewness | 0.8644 | 0.8348 | 0.8205 |
| Gradient Vectors Kurtosis | 0.6549 | 0.6327 | 0.6228 |

Base on Table 5.6, we can see that all of the seven features have AUC significantly greater than random guess (AUC=0.5). Among all of them, two features:

thickness skewness and gradient vector variance have very high specificity together with sensitivity, for both of them have an $AUC > 0.9$. Meanwhile, the gradient vector kurtosis has the worst performance in this test, with an $AUC = 0.6549$. One might question that if you have a feature that has discriminative power of $AUC$ over 0.95, one should consider using *only* this feature to classify all the subjects. However, in reality, one feature might work well in a given subject but fail terribly in another due to the heterogeneity of the FCD lesions. Another example of feature analysis on 50000 randomly selected cubes of subject #1 using ROC curves is presented in Table 5.7. One can see that the performance of individual feature is very different from Table 5.6. The gradient vector kurtosis feature, which has the worst performance for subject #1, has the best performance with $AUC = 0.7447$ in subject #1. Also, the absolute gradient feature, which has an $AUC = 0.8753$ in for subject #50, can only produce $AUC = 0.5035$ for this subject. In clinical practice, it is often hard to know before hand which feature will perform better in a particular subject. In this study, we use all the 7 features for the subsequent analysis. The ROC statics for the entire data set is summarized in Table 5.8, which we can see the performance is much worse than the performance we see in Table 5.6. The ones have better performances are the thickness, thickness kurtosis features, and the gradient vector variances features, with AUC around 0.65. Note that even though AUC has been widely used in the medical community to evaluate the diagnosis testing, the value has no clinically relevant meaning. Also, a lot of the area is coming from the range of large false positive or false negative values, which may never be considered in a medical application. Here, we provide these ROC curves and AUC statistics only to demonstrate the effectiveness of our features.

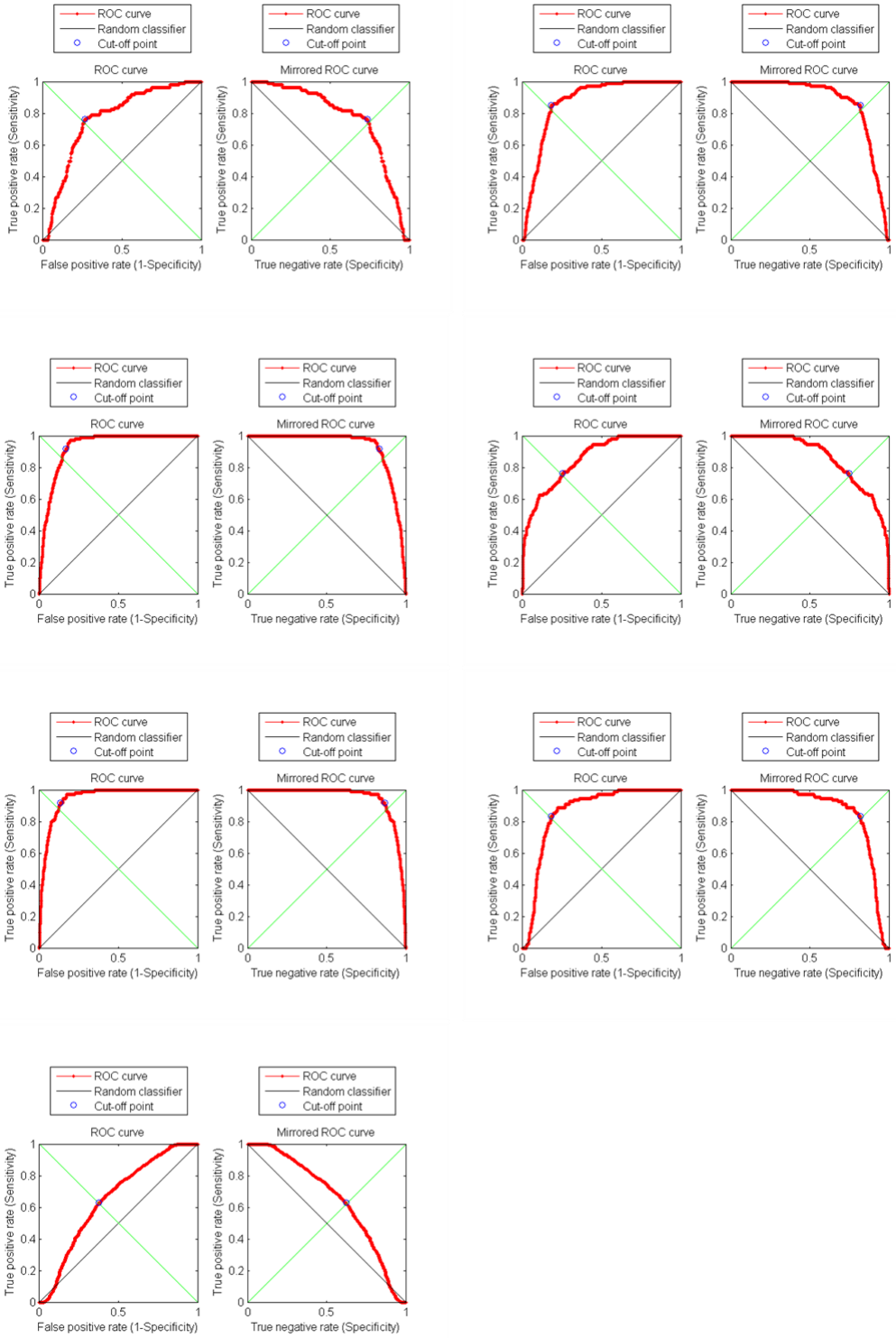


Figure 5.5: The feature ROC curves of subject #50. (top left)thickness (top right)gradient (2nd row left)gradient vector variance (2nd row right) thickness skewness (3rd row left) thickness kurtosis (3rd row right) gradient skewness (4th row) gradient kurtosis

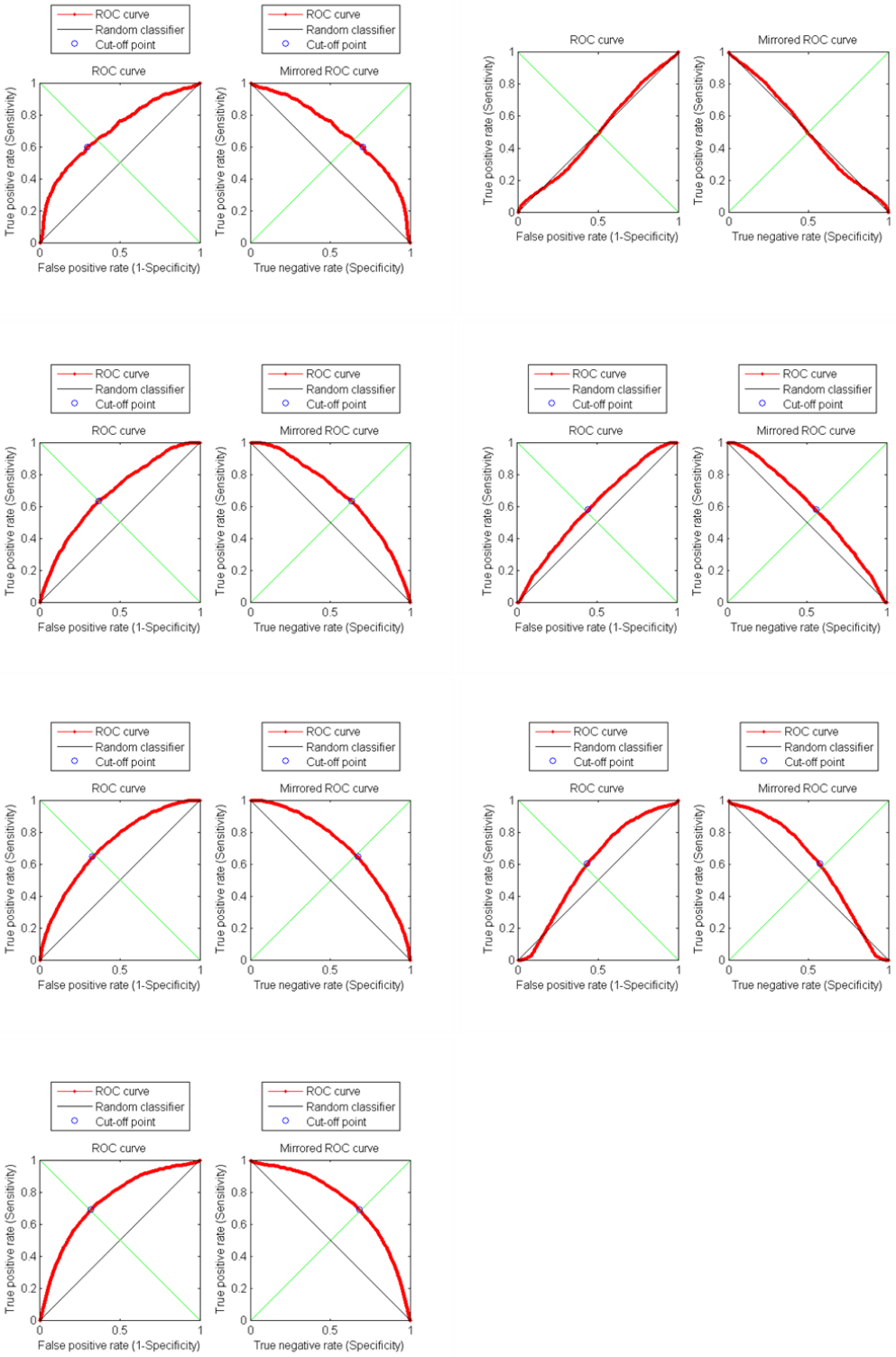


Figure 5.6: The feature ROC curves of subject #1. (top left)thickness (top right)gradient (2nd row left)gradient vector variance (2nd row right) thickness skewness (3rd row left) thickness kurtosis (3rd row right) gradient skewness (4th row) gradient kurtosis

Table 5.7: ROC statistics for proposed features for subject # 1

| Feature | AUC | sensitivity | specificity |
|---|---|---|---|
| Cortical Thickness (cm) | 0.7046 | 0.5985 | 0.7049 |
| Thickness Skewness | 0.6038 | 0.5811 | 0.5559 |
| Thickness Kurtosis | 0.7213 | 0.6419 | 0.6818 |
| Absolute Gradient | 0.5035 | N/A | N/A |
| Gradient Vectors Variance | 0.6842 | 0.6355 | 0.6349 |
| Gradient Vectors Skewness | 0.5976 | 0.6028 | 0.5722 |
| Gradient Vectors Kurtosis | 0.7447 | 0.6942 | 0.6818 |

Table 5.8: ROC statistics for proposed features for all subjects

| Feature | AUC | sensitivity | specificity |
|---|---|---|---|
| Cortical Thickness (cm) | 0.6361 | 0.6763 | 0.6242 |
| Thickness Skewness | 0.6001 | 0.5754 | 0.6005 |
| Thickness Kurtosis | 0.6654 | 0.6242 | 0.63 |
| Absolute Gradient | 0.5970 | 0.6848 | 0.5049 |
| Gradient Vectors Variance | 0.6680 | 0.6141 | 0.6559 |
| Gradient Vectors Skewness | 0.6342 | 0.6181 | 0.5997 |
| Gradient Vectors Kurtosis | 0.5955 | 0.6779 | 0.4802 |

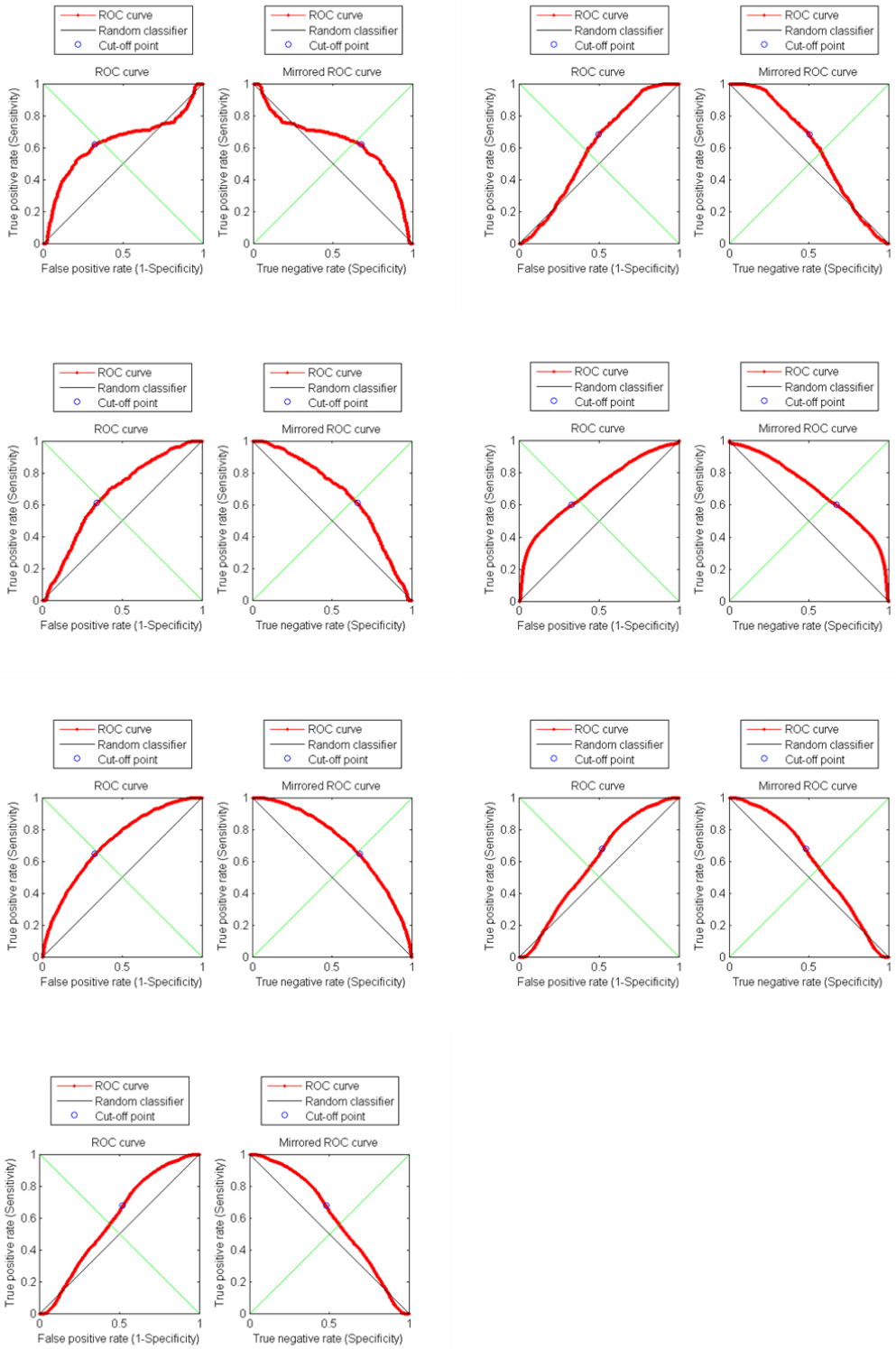


Figure 5.7: The feature ROC curves for all subjects. (top left)thickness (top right)gradient (2nd row left)gradient vector variance (2nd row right) thickness skewness (3rd row left) thickness kurtosis (3rd row right) gradient skewness (4th row) gradient kurtosis

## 5.3 Summary

In this chapter, several statistical tools are used to demonstrate the effectiveness of the proposed T1 volume-based features. From the statistical results, we can conclude that the proposed features are able to provide prediction power for classifying healthy tissues from FCD lesions. The usefulness of these extracted features is very important for the overall performance of the CAD systems. In the next chapter, we will evaluate the CAD detection and segmentation performance for using T1 volume-based, T1 cluster-based features for FCD lesion detection, and the MAP-MRF framework for T2 FCD lesion segmentation.

# Chapter 6

# Experimental Results

In this chapter, we demonstrate classification/segmentation results on the two proposed CAD systems for both T1 and T2 FCD lesion detection and segmentation. In Section 6.1, we present the results on volume-based features for T1-weighted MRI with a Naive Bayes Classifier. The binary classification performance, and the posterior probability of the Naive Bayes modeling is studied. In Section 6.2.1, the result of classifying cluster-wise differential features using SVM classifier on T1 weighted is presented. Section 6.3.1 the segmentation results on the segmentation of FCD lesion in T2 weighted MRI using MRF with embedded T1 location prior are presented. Finally, in Section 6.3.2, the segmentation results on T1 and T2 are compared and discussed.

## 6.1 Naive Bayes Modeling and Classification for T1 Volume-Based Features

### 6.1.1 Cross-Validation

One of the most important issues with the proposal of a new CAD algorithm is the process of validation. One needs to verify the propose method with a dataset that is large enough in order to conclude robustness of the method. In

this study, the classification results are all achieved by *Leave-One-Out-Cross-Validation*(LOOCV) algorithm using a total of 51 subjects as described in Section 5. For any given dataset $\{\{\mathbf{x_i}, y_i\}$ where $\mathbf{x_i}$ is a feature vector and $y_i$ is the corresponding class label. The dataset $\{\{\mathbf{x_i}, y_i\}$ comes from a total of $N$ subjects, the notion $x_i \in l$ with $l \in [1, N]$ means that the feature vector $x_i$ comes from subject $l$. The algorithm of LOOCV is summarized as follows.

**Algorithm: LOOCV**

1. Input labeled examples $\{\mathbf{x_i}, y_i\}$
2. For $l = 1$ to $N$ do
   a. Train a classifier $h$ on $\{\mathbf{x_i}, y_i\}$ $\{\mathbf{x_i}, y_i\} \in j$
   b. Classify $x_i \in l$ by $h(\mathbf{x_i} \in \mathbf{l})$
3. Calculate the corresponding classification performance metrics for $h(\mathbf{x_i} \in \mathbf{l})$
4. Average all the performance for each $h(\mathbf{x_i} \in \mathbf{l})$

The LOOCV provides an efficient use of available data when the data is not large enough to divide it into training and testing set. However, LOOCV can be very computationally expensive since it requires one iteration for each data. An alternative of the LOOCV method is a Leave-K-Out-Cross-Validation (LKOCV), which it means leaving $k$ subjects from the training set and the evaluation is based on testing the $k$ subjects on the trained classifier.

### 6.1.2 Posterior Probability Analysis

For a 2-class Naive Bayes Classifier with equal prior probability assumption $P(H) = P(L) = \frac{1}{2}$, a cube with feature vector $\mathbf{F}$ will be classified as lesional as long as its posterior probability $P(\mathbf{F}|L) > P(\mathbf{F}|H)$. In other words, if $(\mathbf{F}|L) > 0.5$, the cube would be classified as lesional. In this section, we investigates this poster probability for the Naive Bayes Classifier. We are not only interested in knowing that a

cube is being classified as lesional, but also we would like to see how "confident" the classifier consider a cube as lesional and wether it correlates with the actual lesion locations. Determining the posterior probability can also help eliminate the number of false positives, see Fig 6.1 for example.

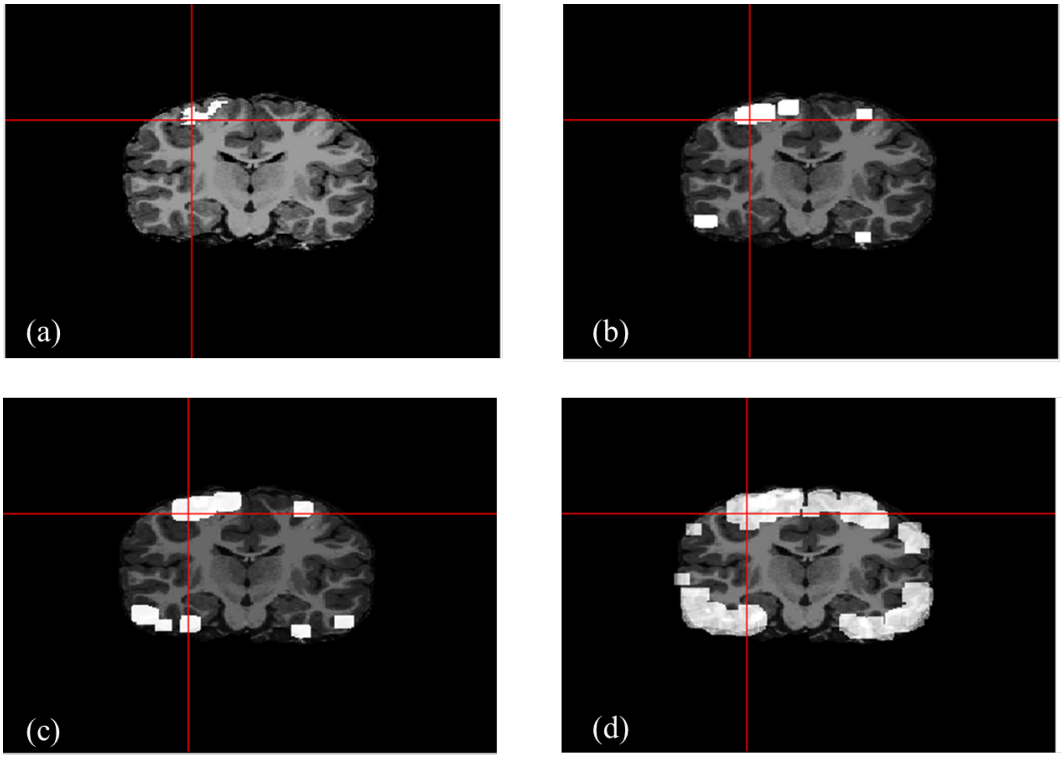


Figure 6.1: Example result: FCD lesion predictions based on different posterior probability $P(L|F)$ of subject #1.(a) Actual lesion (b) lesion detected with $P(L|F) = 0.98$ (c) lesion detected with $P(L|F) = 0.90$ (d) Lesion detected with $P(L|F) = 0.50$

To analyze the results, we look at the posterior probability of the Naive Bayes model for the actual lesions and compare it with the posterior probability of the healthy ones. Assume we have a set of feature vectors $F_h$ which are in the healthy class, and another set of feature vectors $F_l$ which are labeled lesional. We are interested in knowing how the Naive Bayes probabilistic model think of these two sets. Assume $P(L|F_l)$ is the average confidence of the features being

lesional from the probabilistic model with given feature set $F_l$. Similarly, $P(L|F_h)$ is the classifier's average confidence of the feature sets $F_h$ being lesional. Ideally, we hope that $P(L|F_l) \gg P(L|F_h)$ and $P(L|F_l) \approx 1$. To further test the need for using all seven features, we provide the results with using only 3 features (thickness, absolute gradient, gradient vector variance), 5 features (plus thickness skewness and kurtosis), and all the proposed 7 features. Table 6.1

summarizes the results using Naive Bayes probabilistic model trained with different number of features (3, 5 and 7) and using the same LOOCV setting as we discussed the previous section.

From Table 6.1, we can see the use of additional distribution statistic features greatly improves probabilistic model's ability to differentiate lesional and healthy features. When using only 3 features, the overall confidence of considering lesional features as lesion $P(L|F_l)$ is only 0.4999, just slightly greater than considering the healthy feature sets $P(L|F_h) = 0.4436$. For the 5-feature case, the number improve to 0.5188 and 0.4355. While using 7 features give us the best performance of 0.5260 and 0.4325. The performance does not only improve averagely but also increase in each individual patient. With the additional features, the probabilistic model gives more accurate estimates of the features are lesional or not and the results are consistent for each individual patient.

Besides knowing overall confidence of the lesional area, we are also interested in analyzing the areas with relatively high probability (e.g. more than 0.90) of being lesional. It is desirable if the actual lesional area are a much given higher probability from the probabilistic output. One can use a high threshold of eliminating false positive or less lesion-probable regions without taking the risk of missing actual lesions. Table 6.2 summarizes the results on the actual lesions detected using a probability 0.90 threshold as "hit" or "detected" for lesions. The results are the output of the 7-feature Naive Bayes Probabilistic Model.

One can see from Table 6.2 that if using 0.9 as a confidence threshold, all of the

Table 6.1: Posterior Probability by Naive Bayes Probabilistic Model

| Patient ID | 3 features | | 5 features | | 7 features | |
|---|---|---|---|---|---|---|
| | $P(L\|F_l)$ | $P(L\|F_h)$ | $P(L\|F_l)$ | $P(L\|F_h)$ | $P(L\|F_l)$ | $P(L\|F_h)$ |
| 1 | 0.5233 | 0.4397 | 0.7019 | 0.4238 | 0.7224 | 0.4147 |
| 2 | 0.4775 | 0.4396 | 0.6276 | 0.4346 | 0.6400 | 0.4284 |
| 3 | 0.4507 | 0.4419 | 0.4637 | 0.4343 | 0.4751 | 0.4323 |
| 4 | 0.4494 | 0.4403 | 0.6325 | 0.4360 | 0.6327 | 0.4320 |
| 5 | 0.4606 | 0.4388 | 0.6285 | 0.4465 | 0.6407 | 0.4486 |
| 6 | 0.4667 | 0.4495 | 0.4730 | 0.4403 | 0.4732 | 0.4359 |
| 7 | 0.6975 | 0.4450 | 0.7048 | 0.4322 | 0.6843 | 0.4304 |
| 8 | 0.3586 | 0.4395 | 0.3500 | 0.4257 | 0.3780 | 0.4251 |
| 9 | 0.6391 | 0.4468 | 0.6067 | 0.4339 | 0.6139 | 0.4347 |
| 10 | 0.4732 | 0.4551 | 0.5680 | 0.4582 | 0.5889 | 0.4654 |
| 11 | 0.6224 | 0.4134 | 0.6750 | 0.4299 | 0.6947 | 0.4132 |
| 12 | 0.5358 | 0.4534 | 0.6120 | 0.4423 | 0.6340 | 0.4219 |
| 13 | 0.5129 | 0.4420 | 0.6911 | 0.4355 | 0.7014 | 0.4598 |
| 14 | 0.4758 | 0.4153 | 0.4620 | 0.4027 | 0.5341 | 0.4152 |
| 15 | 0.3475 | 0.4357 | 0.3399 | 0.4146 | 0.3679 | 0.4140 |
| 16 | 0.4383 | 0.4504 | 0.6214 | 0.4471 | 0.6216 | 0.4431 |
| 17 | 0.5459 | 0.4315 | 0.6053 | 0.4218 | 0.6458 | 0.4119 |
| 18 | 0.4912 | 0.4135 | 0.5214 | 0.4518 | 0.5289 | 0.4577 |
| 19 | 0.6113 | 0.4023 | 0.6649 | 0.4185 | 0.6823 | 0.4012 |
| 20 | 0.5470 | 0.4588 | 0.6976 | 0.4300 | 0.7259 | 0.4166 |
| 21 | 0.4549 | 0.4212 | 0.4738 | 0.4415 | 0.4720 | 0.4380 |
| 22 | 0.5315 | 0.4328 | 0.5647 | 0.4139 | 0.5771 | 0.4209 |
| 23 | 0.6774 | 0.4140 | 0.6882 | 0.4214 | 0.7015 | 0.4125 |
| 24 | 0.5813 | 0.4776 | 0.6045 | 0.4521 | 0.6097 | 0.4498 |
| 25 | 0.5056 | 0.4353 | 0.5158 | 0.4334 | 0.5469 | 0.5140 |
| 26 | 0.4910 | 0.4219 | 0.4893 | 0.4177 | 0.4955 | 0.4761 |
| 27 | 0.4890 | 0.4001 | 0.5009 | 0.4134 | 0.5128 | 0.4298 |
| 28 | 0.5224 | 0.4332 | 0.5243 | 0.4134 | 0.5909 | 0.4172 |
| 29 | 0.5309 | 0.4527 | 0. 5587 | 0.4290 | 0.5439 | 0.4035 |
| 30 | 0.4881 | 0.4278 | 0. 5193 | 0.4133 | 0.5485 | 0.4198 |
| 31 | 0.5455 | 0.4114 | 0.6006 | 0.4288 | 0.6297 | 0.4255 |
| 32 | 0.6014 | 0.4013 | 0.6543 | 0.3909 | 0.7013 | 0.3888 |
| 33 | 0.5358 | 0.4250 | 0. 5218 | 0.4234 | 0.5202 | 0.4345 |
| 34 | 0.6419 | 0.4512 | 0.6508 | 0.4557 | 0.6774 | 0.4683 |
| 35 | 0.4962 | 0.4312 | 0.5015 | 0.4219 | 0.5238 | 0.4315 |
| 36 | 0.4940 | 0.4719 | 0.4990 | 0.4824 | 0.5193 | 0.4724 |
| 37 | 0.5313 | 0.4866 | 0.5884 | 0.4353 | 0.6003 | 0.4211 |
| 38 | 0.4848 | 0.4450 | 0.4935 | 0.4438 | 0.4990 | 0.4358 |
| 39 | 0.4958 | 0.4184 | 0.5311 | 0.4586 | 0.5331 | 0.4881 |
| 40 | 0.6283 | 0.4301 | 0.6507 | 0.4212 | 0.6994 | 0.4540 |
| 41 | 0.5110 | 0.4728 | 0.5235 | 0.4594 | 0.5390 | 0.4426 |
| 42 | 0.6238 | 0.4903 | 0.6538 | 0.4812 | 0.6633 | 0.4874 |
| 43 | 0.3940 | 0.4212 | 0.4150 | 0.4238 | 0.4884 | 0.4221 |
| 44 | 0.4345 | 0.4111 | 0.4629 | 0.4182 | 0.4909 | 0.4269 |
| 45 | 0.4323 | 0.4296 | 0.4644 | 0.4189 | 0.4767 | 0.4057 |
| 46 | 0.7320 | 0.4522 | 0.7485 | 0.4438 | 0.7593 | 0.4131 |
| 47 | 0.3545 | 0.4497 | 0.4206 | 0.4625 | 0.4544 | 0.4597 |
| 48 | 0.6107 | 0.4325 | 0.6296 | 0.4228 | 0.6501 | 0.4204 |
| 49 | 0.5134 | 0.4093 | 0.5228 | 0.4038 | 0.5445 | 0.4021 |
| 50 | 0.7014 | 0.4272 | 0.7273 | 0.4413 | 0.7535 | 0.4250 |
| 51 | 0.4563 | 0.4544 | 0.4899 | 0.4953 | 0.4961 | 0.4711 |

lesions are detected partially. The least hit rate is 33% and the highest reaches up to 96%. The results suggest that the Naive Bayes probabilistic model provide us a reasonable estimation of the probable lesion location based on the 3D volume-base features on the T1 weighted MRI.

### 6.1.3 Comparison with Antel's Features

The classification results are derived using the Naive Bayes probabilistic modeling described in Section 2.4.1 with MAP decision rule defined in eq.(2.2). Before classification, each of the feature vectors was first normalized to negate the variable's effect on the classification results. The Naive Bayes classifier was trained by using Leave-One-Out Cross Validation (LOOCV) to classify the feature vectors. In other words, the training set size is $n-1$ subjectwise when there are $n$ subjects available and the performance is evaluated based on averaging the performance from running a number of subsets. Each subject's feature vectors constructed from $8 \times 8 \times 8$ voxel cubes were classified using the classifier trained by feature vectors obtained from the other subjects. The performance is evaluated based on *detection rate* and *false positive rate*. Note that both evaluations are computed in the sense of feature vectors instead of subjects. In a subjectwise manner, the classifier identified lesions in 51/51 subjects, which is a subjectwise sensitivity of 100%.

For a binary classification task using MAP rule, the label $C \in \{H, L\}$ is assigned if the posterior probability $P(C|F_1, F_2, ..., F_n) > 0.5$. Note that other thresholds could be chosen for the detection task. A more conservative threshold like 0.5 would potentially allow more false positives. However, in our setting, a 0.5 probability is reasonable in the sense that the healthy and lesional tissues are sometimes very similar, so we select a more conservative threshold. From Eq.(**??**), we know that other than the modeled distributions $P(F_i|C)$, the posterior probability is also determined by the class priors $P(C)$. A high $P(L)$ will provide

Table 6.2: Number of lesional cubes with more than 0.9 of posterior probability

| Patient ID | lesional cubes | of cubes $P(L\|f_l) > 0.9$ | hit ratio |
|---|---|---|---|
| 1 | 1287 | 1236 | 96.04% |
| 2 | 577 | 416 | 72.10% |
| 3 | 450 | 443 | 98.44% |
| 4 | 3688 | 3511 | 95.20% |
| 5 | 5239 | 4480 | 52.39% |
| 6 | 375 | 113 | 30.10% |
| 7 | 434 | 395 | 91.01% |
| 8 | 1030 | 954 | 92.62% |
| 9 | 2455 | 2437 | 99.27% |
| 10 | 3429 | 1339 | 39.05% |
| 11 | 1621 | 816 | 50.34% |
| 12 | 2004 | 1414 | 70.56% |
| 13 | 1234 | 1005 | 81.44% |
| 14 | 2896 | 2417 | 83.46% |
| 15 | 1834 | 982 | 53.54% |
| 16 | 7254 | 5999 | 72.54% |
| 17 | 118 | 42 | 35.59% |
| 18 | 4230 | 4188 | 99.01% |
| 19 | 7245 | 2391 | 33.00% |
| 20 | 1921 | 653 | 33.99% |
| 21 | 987 | 970 | 98.28% |
| 22 | 2407 | 2016 | 83.76% |
| 23 | 230 | 175 | 76.30% |
| 24 | 982 | 711 | 72.40% |
| 25 | 2396 | 2268 | 94.66% |
| 26 | 1002 | 932 | 93.01% |
| 27 | 5317 | 2019 | 37.97% |
| 28 | 532 | 425 | 77.26% |
| 29 | 911 | 850 | 93.30% |
| 30 | 2063 | 1397 | 67.72% |
| 31 | 1216 | 859 | 70.64% |
| 32 | 209 | 158 | 75.60% |
| 33 | 410 | 173 | 42.20% |
| 34 | 6456 | 3584 | 55.69% |
| 35 | 1442 | 1442 | 100% |
| 36 | 1777 | 1688 | 94.99% |
| 37 | 2817 | 2100 | 74.55% |
| 38 | 2376 | 2350 | 98.91% |
| 39 | 2817 | 1264 | 44.87% |
| 40 | 54 | 40 | 74.07% |
| 41 | 4903 | 1705 | 34.77% |
| 42 | 5155 | 5094 | 98.82% |
| 43 | 641 | 600 | 93.60% |
| 44 | 1382 | 1128 | 81.62% |
| 45 | 114 | 65 | 57.02% |
| 46 | 5553 | 5240 | 94.36% |
| 47 | 2014 | 1997 | 99.65% |
| 48 | 1139 | 621 | 54.52% |
| 49 | 11011 | 6502 | 59.03% |
| 50 | 218 | 189 | 86.70% |
| 51 | 1258 | 1237 | 98.33% |

a high detection rate meanwhile introduce a high number of false positives. In our study, we calculate the detection rates and false positives for different class prior for $P(L)$ from 0 to 1 with 0.05 intervals.

In order to see if the additional distributional features indeed increase discriminating power and enhance classification performance, we ran the experiments over subsets of features by using 3 (cortical thickness, absolute gradient, variance of orientation), 5 (3 features plus skewness and kurtosis of thickness) and 7 (5 features plus skewness and kurtosis of orientation). The classification results obtained by training different numbers of features are plotted in Fig.3. The plot is generated by changing the class priors $P(L)$ and each dot in the figure represents the corresponding detection and false positive rates. From Fig.3 we can see that the classifier trained by using 7 features has the best performance among the three. With equal priors $P(L) = P(H) = 0.5$ assumption, there is a 62.49% detection rate and 19.31% false positive rate.

We also compare our proposed features with Antel's features proposed in [36] using cortical thickness, gradient, and relative intensity as features calculated on a voxel-wise basis. The performance results in Fig.4 indicate a better discrimination is achieved by using the proposed features. The marked points on Fig.4 are the performances from the two classifiers both using equal class priors. One can observe from Fig.4 that using Antel's features, while yielding a high detection rate even with low $P(L)$, there is a high number of false positives. This could be due to the fact that Antel's features are unable to differentiate GM from lesions successfully, as discussed in [7].

## 6.2 T1 Cluster-based Feature Analysis

### 6.2.1 Support Vector Machine Classification

For volume-based classification, we set the threshold of Naive Bayes posterior probability of being lesions to 0.95 and detected FCD lesions in 51/51 subjects.

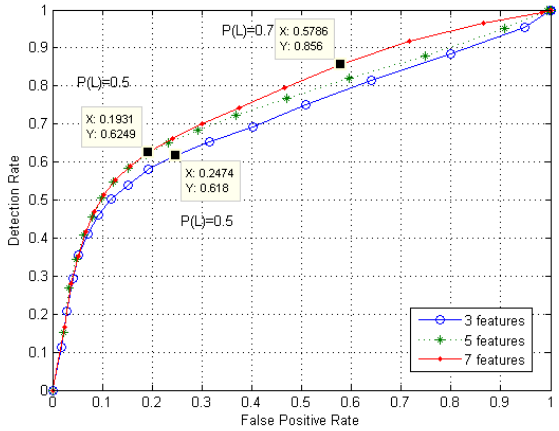


Figure 6.2: The classification performance of Naive Bayes classifier trained by 3, 5 and 7 features

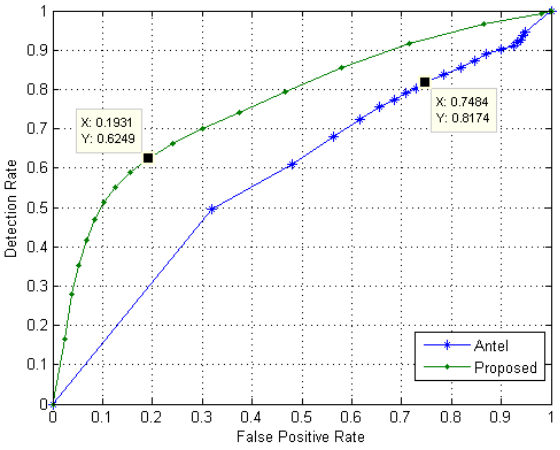


Figure 6.3: The classification performance of Naive Bayes classifier trained by proposed features and Antel's features

However, there are several false positive clusters present in the subjects. A total of 402 clusters are marked as lesional candidates, averaging 7.88 clusters per subject. Among all the lesional candidates, only 95/402 are actually lesions. The size of the candidate clusters is ranging from 512 voxels to 3460 voxels. In this study, there is no pre-defined thresholds set for the minimum cluster size as FCD lesions often varied greatly in size.

In the feature construction steps as described in Section 3.4, we constructed a feature vector for each cluster, each feature vector is consist of a total of 13 features. We first test the proposed cluster-based features using the whole dataset for both training and testing on a SVM classifier with 2nd-order polynomial kernel. In this setting, all 95 actual lesions were successfully detected with no false positives (FP). Note that if we train and test with only the 7 original absolute features, the result shows significantly lower performance where only 8 actual lesions is detected.

In a more realistic setting, we used 1/3 of the subjects for training and other 2/3 of subjects for testing. The experimental results are presented in Table 6.3. From Table 6.3, we can see that the using cluster-wise classification step has greatly reduce the number of FP. Note that the classifier has a preference toward healthy label due to the fact that we have imbalance class labels in our training sets. (95 in lesional class, 307 in healthy class) Table 6.3 also compares the performance of using only the absolute features and with additional differential features. The results show that using only the absolute features could not achieve good classification results, which is not surprising since the lesional candidates have all been considered as lesional using the same set of features in the previous step. On the other hand, with the use of differential features, a significant increase in discriminative performance is achieved. The classifier is able to identify 87% of the actual lesions with only 3.94 FP per case. The results indicates that the differential features are significantly more discriminative than absolute ones.

Table 6.3: Classification result of cluster-wise features on T1-weighted MRI

| | Initial | Absolute features | With differential features |
|---|---|---|---|
| Detection Rate | 100% | 24% | 87% |
| FP per case | 7.88 | 4.17 | 3.94 |

## 6.3 Experimental Results on T2

### 6.3.1 Segmentation Results

The proposed MRF segmentation algorithm was experimented on 10 MRI brain volumes with FCD lesions. The accuracy of our results are computed using the traditional similarity index or the $\kappa$ statistic which is defined as

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

where A is the set of voxels that is labeled as lesional and B is the set of voxels labeled as lesion by the proposed algorithm. Traditionally, a value for $\kappa$ statistic that is more than 0.7 represents a very good segmentation in the lesion segmentation field [93], with state-of-the art method like Van Leemput et al. [67] reported around $0.47 - 0.51$ and a more recent work by Shiee et al. [64] with 0.677 on segmenting Multiple Sclerosis (MS) lesions. However, comparing the $\kappa$ statistics of different data sets for different diseases is of little value. One should only compare $\kappa$ statistics when two algorithms are using the same data set and training criterion. Although the $\kappa$ statistic is a reasonable measure of similarity between the algorithm and the experts, it provides no information about the kind of errors generated by the algorithm. Therefore, here we also present the false positive (FP) rate and the false negative (FN) rates, so that we can measure the what is the cause of inaccuracy in our algorithm. The classification results of 10 FCD subjects using $\kappa$ statistic, FPs and FNs are summarized in Table 6.4.

Table 6.4: Classification result of MRF modeling on T2-weighted MRI

| subject ID | $\kappa$ statistic | FP | FN |
|---|---|---|---|
| 1 | 0.7245 | 0.0022 | 0.0934 |
| 2 | 0.6630 | 0.0011 | 0.1288 |
| 3 | 0.6884 | 0.0035 | 0.1350 |
| 4 | 0.5992 | 0.0018 | 0.1627 |
| 5 | 0.7005 | 0.0049 | 0.0845 |
| 6 | 0.6259 | 0.0036 | 0.1041 |
| 7 | 0.5147 | 0.0028 | 0.1223 |
| 8 | 0.6334 | 0.0041 | 0.0612 |
| 9 | 0.6255 | 0.0054 | 0.0855 |
| 10 | 0.7140 | 0.0016 | 0.0661 |
| mean | 0.6487 | 0.0031 | 0.1044 |

From Table 6.4, we can see that the average $\kappa$ statistic for our algorithm is 0.6487, which is comparable to the state-of-the-art methods for the MS lesions segmentation. Also note that the $\kappa$ statistics compare the lesions marked by the algorithm with the lesions outline by a single trained human. It is reported that it is often a lack of consensus even among trained experts for the task of FCD lesion segmentations [16]. The FP values are generally low with the value varies between 0.0018 to 0.0054, and the FN values are at most 16%. The low FN values indicate that the algorithm is able to capture most lesions with a high coverage rate around 90%.

### 6.3.2 Performance analysis of Location Prior

In this section, we investigate the classification performance of our MRF model with the addition the location prior. For the investigation purposes, here the MRF appearance model is not learnt by EM algorithms and chosen as various thresholds instead. Assume the maximum intensity of the brain is $I_{max}$, we define various threshold $t_i$'s from $0.1 - 0.9$ times $I_max$. Everything above the thresholds we assume it to be belongs to the lesional distribution. By alternating the value $t_i$, we will get a different sensitivity and specificity for each $t_i$. The results can be

summarized using ROC curves and observe the AUC statistics, which we can see in Fig 6.4, an example of the segmentation result of subject #47 is provided. The segmentation performance ROC curves are generated by using MRF modeling in intensity and with location prior as describe in Section 4.4. As we can see in Fig 6.4, using the location prior greatly increase the prediction accuracy from $AUC = 0.7490$ to $AUC = 0.8241$.

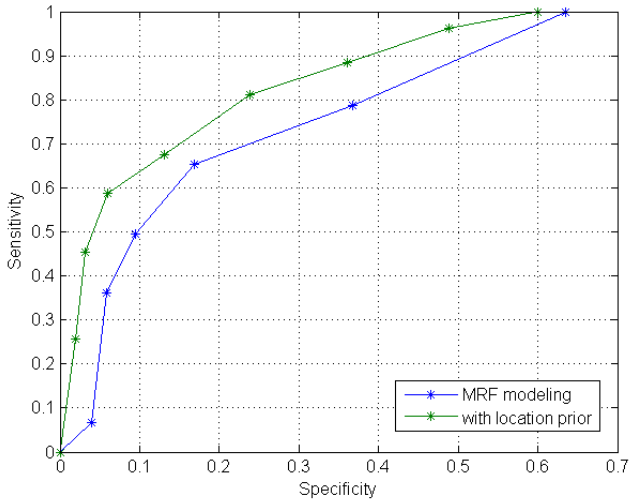


Figure 6.4: The segmentation performance of MRF intensity segmentation alone and with the location prior of subject #47. The green line is MRF with appearance and location prior AUC =0.8241, with only the apperance prior AUC = 0.7490

## 6.4 Summary

In this chapter, we provided the classification performance for T1 volume-based and cluster-based detection tasks using Naive Bayes Classifier and SVM. For T2 FCD lesion segmentation task, the MAP-MAR framework provided a comparable results with the state-of-the art methods and the use of location prior gave better prediction accuracy. Experimental results demonstrate the proposed methods are able to identify lesion with a very high detection rate 100% and 87% for

volume-based and cluster-based T1 FCD lesions and 100% for T2 FCD lesions. The improvements we made to the traditional systems, such as the volume-based and cluster-based features outperforms the performance of Antel's system; also the additional location prior increases the detection accuracy of the traditional MAP-MRF framework. In the next chapter, the thesis is discussed and concluded.

# Chapter 7

# Conclusion and Discussion

In this chapter, some future work on this related topic is suggested in Section 7.1 and the thesis is summarized in Section 7.2.

## 7.1 Future Work

Several possibilities involve for related future work on the topic. First of all, when we try to perform the mutual information registration of T1 and T2 images of the subjects with FCD lesions, the registration results are sometimes not accurate. We suspect that it may be due to the presence of FCD lesions could affect the distributions of the tissue intensity therefore reduce the accuracy of the mutual information registration. One possible investigation can be done for this problem is to perform the registration without considering the lesional area. Another potentially interesting work can be incorporation of other priors into the MAP-MRF framework described in Section 4.4, some of the features in T1 can also be valuable information to assist segmentation in T2.

Also, combining different imaging modalities could also be possible to provide more accurate results for FCD lesion analysis and detection. For example, traditional EEG signal or functional neuroimaging modalities such as positron

emission tomography (PET) and single-photon emission computerized tomography (SPECT) can provide excellent insights to the functional state of the brain. Adding such information can increase the dimension of study of the FCD lesion detection.

The selections of parameters and classifiers are not optimal in this thesis. Tuning parameters, such as the volume of the cube size and weights on the terms in the modification of the SVD, could be better learned by a cross-validation if one wants to achieve optimal solution. Also, there is some additional work that can be done in choosing the classifiers, there are a lot of appropriate machine learning algorithms that have not been fully considered due to time constrains. Also, the proposed methods have not been tested on healthy controls, which can also be investigated in the future.

## 7.2 Conclusion

Focal Cortical Dysplasia (FCD) is the most common cause of medically refractory epilepsy. Detection and segmentation of such lesions on MRI is a challenging task even for trained experts. In this thesis, we proposed two CAD systems for FCD lesions detection and segmentation on T1 and T2 FLAIR weighted MRI. For the first CAD system, the problem of detection of FCD lesions on T1 weighted MRI is explored. Base on previous findings, the three main characteristics of FCD lesion on T1 weighted MRI are well know, however, the quantitative metrics to characterize such prior knowledge still have their limitations. In this thesis, we proposed using a set of volume-based features. The advantage of volume-based analysis is the ability to process and evaluate a certain spatial neighbored at a time, where potentially more information can be processed and represented. Therefore, the volume-based analysis can be an improvement from the traditional voxel-based analysis. Besides the volume-based settings, the contribution of this thesis is proposed using probabilistic-based methods for the cortical thickness

measurements. Previous methods require the surface of the cortexes to be predefined, which is very difficult for images with FCD lesions.

After the volume-based analysis, a subsequent cluster-based differential features is proposed to reduce the number of false positives from the previous classification using volume-based features and the Naive Bayes Classifier. Differential features are used to compare candidate lesions with other part of the brain within the same subject. From the experimental results, we see that using this cluster-wise feature and a SVM classifier greatly reduce the number of false positives. Two machine learning methods are used here in this thesis for classifications. The first one is a generative method called Naive Bayes classifier. Naive Bayes classifier is used in the first stage of the volume-based analysis for several reasons, one is due to the large number of training examples, and also the availability of expert knowledge to incorporate into the classification. The SVM, on the other hand, is used in the second stage, where smaller training samples are needed for training, and also little has been known of the relationship of candidate cluster locations and other locations of within the brain. Therefore, a discriminative method SVM is used. We believe that selecting the appropriate classifiers would greatly reduce computational cost and improve on the accuracy for the classification.

For the second CAD system, we proposed methods for FCD lesions segmentation on T2 weighted MRI. FCD lesions on T2 weighted MRI are often characterized as hyper-intensified. Segmentation of hyper-intensified signals on T2 images has been studied,however, most of the methods proposed are designed for Multiple Sclerosis. Thus far, to our knowledge, there is no automatic segmentation methods designed for FCD lesions specifically. Here we proposed using a popular MAP=MRF model to approach the problem. Gaussian Mixture Model is used to model the intensity value, estimated by the EM algorithm. This model is incorporate into a MRF modeling as the appearance prior. The main contribution for this system is the proposal of the location prior to tailor the algorithm for detecting FCD lesion. The location prior information is based on the thickness measurements on the cortexes of T1 weighted MRI. T1 weighted MRI is known to provide

better contrast of the soft tissues for brain imaging and lesion segmentation on T2 weighted MRI is less challenging. By combining both T2 hyper-intensity information and T1 location information, we integrate the advantages from both modalities to enhance the robustness of the segmentation. Experimental results have shown that using the location prior gives better performance than using only the appearance prior.

To summarize, two CAD systems are proposed in this thesis that is designed for FCD lesion detections. The main contributions are three fold: First of all, we proposed two new strategies for brain tissue analysis for FCD lesion detection on T1 weighted MRI, which is the volume-based analysis and cluster-based differential analysis. Secondly, we designed a MAP-MRF framework with location prior for T2 FCD lesion segmentation, which has not been done in the literature for segmenting FCD lesions. Last but not least, we designed the CAD algorithms specifically to overcome previous limitations for the task of detecting FCD lesions. To our best knowledge, there has not been any CAD system for detecting FCD lesions used in clinical practice till now. We hope that proposal of these two systems can assist physicians achieving better clinical decisions for FCD lesion diagnosis, which is the main purpose of this thesis.

# References


[1] L. Tassi, N. Colombo, R. Garbelli, S. Francione, G. Lo Russo, R. Mai, F. Cardinale, M. Cossu, A. Ferrario, C. Galli, M. Bramerio, A. Citterio, and R. Spreafico. "Focal cortical dysplasia: neuropathological subtypes, EEG, neuroimaging and surgical outcome". *Brain*, 125(4):1719–1732, 2002.

[2] D. C. Taylor, M. A. Falconer, C. J. Bruton, and J. A. N. Corsellis. "Focal dysplasia of the cerebral cortex in epilepsy". *Journal of Neurology, Neurosurgery and Psychiatry*, 34(3):369–387, 1971.

[3] O. Colliot, N. Bernasconi, N. Khalili, S.B. Antel, V. Naessens, and A. Bernasconi. "Individual voxel-based analysis of gray matter in focal cortical dysplasia". *NeuroImage*, 29(1):162 – 171, 2006.

[4] I. Rish. "An empirical study of the naive Bayes classifier". *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, 3(22):4146, 2001.

[5] V. N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York, Inc., New York, NY, USA, 1995.

[6] N. Colombo, L. Tassi, C. Galli, and A. Citterio. "Focal Cortical Dysplasias: MR Imaging, Histopathologic, and Clinical Correlations in Surgically Treated Patients with Epilepsy". *American Journal of Neuroradiology*, 24(4):724–733, 2003.

[7] S. B. Antel and D. et al. Collins. "Automated detection of focal cortical dysplasia lesions using computational models of their MRI characteristics and texture analysis". *NeuroImage*, 19(4):1748 – 1759, 2003.

[8] P. C. Lauterbur. "Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance". *Nature*, 242(5394):190–191, 1973.

[9] E. M. Haacke. *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. J. Wiley & Sons, Hoboken, New Jersey, USA, 1999.

[10] T. Okuda, Y. Korogi, Y. Shigematsu, and T. Sugahara. "Brain Lesions: When Should Fluid-attenuated Inversion-Recovery Sequences Be Used in MR Evaluation". *Radiology*, 212(3):793–798, 1999.

[11] P. Kwan, S. C. Schachter, and M. J. Brodie. "Drug-Resistant Epilepsy". *New England Journal of Medicine*, 365(10):919–926, 2011.

[12] V. Y. Wang, E. F. Chang, and N. M. Barbaro. "Focal cortical dysplasia: a review of pathological features, genetics, and surgical outcome". *Neurosurgical Focus*, 20(1):1–7, 2006.

[13] P. Krsek, B. Maton, P. Jayakar, P. Dean, B. Korman, G. Rey, C. Dunoyer, E. Pacheco-Jacome, G. Morrison, J. Ragheb, H. V. Vinters, T. Resnick, and M. Duchowny. "Incomplete resection of Focal Cortical Dsplasia is the main predictor of poor postsurgical outcome". *Neurology*, 72(3):217–223, 2009.

[14] S. Koh. "Focal Cortical Dysplasia". *Pediatric Epilepsy Case Studies*, 27:201–206, 2008.

[15] H. Urbach, B. Scheffler, and T. Heinrichsmeier. "Focal Cortical Dysplasia of Taylor's Balloon Cell Type: A Clinicopathological Entity with Characteristic Neuroimaging and Histopathological Features, and Favorable Postsurgical Outcome". *Epilepsia*, 43(1):33–40, 2002.

[16] A. J. Barkovich, R. I. Kuzniecky, G. D. Jackson, R. Guerrini, and W. B. Dobyns. "Classification system for malformations of cortical development: Update 2001". *Neurology*, 57(12):2168–2178, 2001.

[17] A. Bernasconi. "Quantitative MR imaging of the neocortex". *Neuroimaging Clinics of North America*, 14(3):425 – 436, 2004.

[18] A. Yagishita, N. Arai, T. Maehara, H. Shimizu, A. M. Tokumaru, and M. Oda. "Focal cortical dysplasia: appearance on MR images". *Radiology*, 203(2):553–559, 1997.

[19] P. Besson, N. Bernasconi, O. Colliot, A. Evans, and A. Bernasconi. "Surface-based texture and morphological analysis detects subtle cortical dysplasia". *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 5241:645–652, 2008.

[20] A. Barkovich, J. Kuzniecky, and I. Ruben. "Neuroimaging of Focal Malformations of Cortical Development". *Journal of Clinical Neurophysiology*, 13(6):481–494, 1996.

[21] B. Gomez-Anson, M. Thom, N. Moran, J. Stevens, and F. Scaravilli. "Imaging and radiological-pathological correlation in histologically proven cases of focal cortical dysplasia and other glial and neuronoglial malformative lesions in adults". *Neuroradiology*, 42:157–167, 2000.

[22] R. Mitchell, S. J. Karlik, D. H. Lee, and A. Fenster. "Automated detection and quantification of multiple sclerosis lesions in MR volumes of the brain". *SPIE, Medical Imaging VI: Image Processing*, 1652(1):99–106, 1992.

[23] M. L. Giger. "Computer-aided diagnosis of breast lesions in medical images". *Computing in Science Engineering*, 2(5):39 –45, September 2000.

[24] Y. Wang, T. Adali, S. Y. Kung, and Z. Szabo. "Quantification and segmentation of brain tissues from MR images: a probabilistic neural network

approach". *Image Processing, IEEE Transactions on*, 7(8):1165 –1181, aug. 1998.

[25] B. Verma and J. Zakos. "A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques". *Information Technology in Biomedicine, IEEE Transactions on*, 5(1):46 –54, mar. 2001.

[26] J. Lasserre. "Hybrids of Generative and Discriminative Methods for Machine Learning". *PhD thesis, University of Cambridge*, March 2008.

[27] C. M. Bishop and J. Lasserre. "Generative or Discriminative? Getting the Best of Both Worlds". *Bayesian Statistics 8*, (8):3–24, 2007.

[28] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[29] A. Agresti. *An introduction to categorical data analysis*. Wiley-Interscience, Hoboken, New Jersey, USA, 2007.

[30] J. Lafferty, A. Mccallum, and F. Pereira. "Conditional Random Fields: Probabilistic models for segmenting and labeling sequence data". *Proc. 18th International Conf. on Machine Learning*, 3(2):282–289, 2001.

[31] I. Ulusoy and C. M. Bishop. "Generative versus discriminative methods for object recognition". 2(2):258–265, 2005.

[32] A. Bernasconi. "Advanced MRI analysis methods for detection of focal cortical dysplasia". *Epileptic Disord.*, 5(2):81–84, 2003.

[33] F. P. G. Bergo and A. X. Falcao. "Fast and automatic curvilinear reformatting of MR images of the brain for diagnosis of dysplastic lesions". *Biomedical Imaging: Nano to Macro, 3rd IEEE International Symposium on*, 1(1):486 –489, April 2006.

[34] J. Kassubek, H. J. Huppertz, J. Spreer, and A. Schulze-Bonhage. "Detection and localization of focal cortical dysplasia by voxel based 3D MRI analysis". *Epilepsia*, 43(1):592–602, 2002.

[35] M. Wilke, J. Kassubek, S. Ziyeh, A. Schulze-Bonhage, and H. J. Huppertz. "Automated detection of gray matter malformations using optimized voxel-based morphometry: a systematic approach". *NeuroImage*, 20(1):330 – 343, 2003.

[36] S. B. Antel, A. Bernasconi, and N. Bernasconi. "Computational models of MRI characteristics of Focal Cortical Dysplasia improve lesion detection". *NeuroImage*, 17(4):1755 – 1760, 2002.

[37] A. Bernasconi, S. B. Antel, D. L. Collins, and N. Bernasconi. "Texture analysis and morphological processing of Magnetic Resonance Imaging assist detection of Focal Cortical Dysplasia in extra-temporal partial epilepsy". *Ann Neurol*, 49(6):770–775, 2001.

[38] H. Huppertz, C. Grimm, S. Fauser, J. Kassubek, I. Mader, and A. Hochmuth. "Enhanced visualization of blurred gray-white matter junctions in focal cortical dysplasia by voxel-based 3D MRI analysis". *Epilepsy Research*, 67(1-2):35–50, 2005.

[39] A. C. Bastos, R. M. Comeau, and F. Andermann. "Diagnosis of subtle focal dysplastic lesion: Curvilinear reformatting from three-dimensional magnetic resonance image". *Annals of Neurology*, 46(1):88–94, 1999.

[40] J. Rajan, K. Kannan, C. Kesavadas, and B. Thomas. "Focal Cortical Dysplasia (FCD) lesion analysis with complex diffusion approach". *Computerized Medical Imaging and Graphics*, 33(7):553 – 558, 2009.

[41] O. Colliot, T. Mansi, N. Bernasconi, V. Naessens, D. Klironomos, and A. Bernasconi. ""Segmentation of focal cortical dysplasia lesions on MRI using level set evolution". *NeuroImage*, 32(4):1621 – 1630, 2006.

[42] F. P. G. Bergo, A. X. Falcao, C. L. Yasuda, and F. Cendes. "FCD segmentation using texture asymmetry of MR-T1 images of the brain". *Biomedical Imaging: From Nano to Macro, 5th IEEE International Symposium on*, pages 424 –427, May 2008.

[43] O. Colliot, T. Mansi, P. Besson, N. Bernasconi, and A. Bernasconi. "Improved segmentation of focal cortical dysplasia lesions on MRI using expansion towards cortical boundaries". *Biomedical Imaging: Nano to Macro, 3rd IEEE International Symposium on*, pages 323 –326, April 2006.

[44] F. G. Woermann, L. F. Samantha, J. K. Matthias, J. Ashburner, and J. S. Duncan. "Voxel-by-voxel comparison of automatically segmented cerebral gray matter–A rater-independent comparison of structural MRI in patients with epilepsy". *NeuroImage*, 10(4):373 – 384, 1999.

[45] C. Loyek, F. G. Woermann, and T. W. Nattkemper. "Detection of Focal Cortical Dysplasia in MRI using textural features". *Proceedings of BVM 2008*, 1(1):432–436, April 2008.

[46] A. Eow. "Quantitative multi-modal analysis of pediatric focal epilepsy". *Master Thesis, Massachusetts Institute of Technology* , 2005.

[47] T. J. Crow, J. Ball, S. R. Bloom, and Brown et al. "Schizophrenia as an anomaly of development of cerebral asymmetry: A Postmortem study and a proposal concerning the genetic basis of the disease". *Arch Gen Psychiatry*, 46(12):1145–1150, 1989.

[48] J. L. Whitwell. "Voxel-Based Morphometry: An automated technique for assessing structural changes in the brain". *The Journal of Neuroscience*, 29(31):9661–9664, 2009.

[49] J. Ashburner and K.J. Friston. "Voxel-based morphometryThe methods". *NeuroImage*, (11):805–821, 2000.

[50] A. Mechelli, C. J. Price, K. J. Friston, and J. Ashburner. "Voxel-Based Morphometry of the human brain: methods and applications". *Current Medical Imaging Reviews*, 1(2), 2005.

[51] S. E. Jones, B. R. Buchbinder, and I. Aharon. "Three-dimensional mapping of cortical thickness using Laplace's equation". *Human Brain Mapping*, 11(1):12–32, 2000.

[52] D. Flanders, H.-B Mryka, and J. Pereverzoff. "Preliminary evaluation of eCognition object-based software for cut block delineation and feature extraction". *Canadian Journal of Remote Sensing*, 29(4):441–452, 2003.

[53] X. Llad, O. Ganiler, A. Oliver, R. Mart, J. Freixenet, L. Valls, J. C. Vilanova, L. Rami-Torrent, and A. Rovira. "Automated detection of multiple sclerosis lesions in serial brain MRI.". *Neuroradiology*, 2011.

[54] J. Grimaud, M. Lai, J. Thrope, P. Adeleine, and L. Wang. "Quantification of MRI lesion load in multiple sclerosis: A comparison of three computer-assisted techniques". *Magnetic Resonance Imaging*, 14(5):495 – 505, 1996.

[55] D. L. Pham, C. Y. Xu, and J. L. Prince. "Current methods in medical image segmentation". *Annual Review Of Biomedical Engineering*, 2:315–337, 2000.

[56] J. Freixenet, X. Munoz, D. Raba, J. Marti, and X. Cufi. "Yet another survey on image segmentation: Region and boundary information integration". *Proceedings of the third European conference on Computer Vision (ECCV)*, 2352:21–25, 2002.

[57] M. Filippi, M. A. Horsfield, P. S. Tofts, F. Barkhof, A. J. Thompson, and D. H. Miller. "Quantitative assessment of MRI lesion load in monitoring the evolution of multiple sclerosis". *Brain*, 118(6):1601–1612, 1995.

[58] P. K. Sahoo, S. Soltani, and Wong A. K. C. "A survey of thresholding techniques". *Computer Vision, Graphics, and Image Processing*, 41(2):233 – 260, 1988.

[59] S. Warfield, D. Joachim, Z. Joachim, C. R. G. Guttmann, W. M. Wells, G. J. Ettinger, J. Hiller, and R. Kikinis. "Laboratory investigation: automatic identification of gray matter structures from MRI to improve the segmentation of white matter lesions". *Computer Aided Surgery*, 1(6):326–338, 1995.

[60] Z. Pan and J. Lu. "A Bayes-based region-growing algorithm for medical image segmentation". *Computing in Science and Engineering*, 9:32–38, 2007.

[61] Z. Lao, D. Shen, D. Liu, F. J. Abbas, E. R. Melhem, L. J. Launer, R. N. Bryan, and C. Davatzikos. "Computer-assisted segmentation of white matter lesions in 3D MR images using Support Vector Machine". *Academic Radiology*, 15(3):300 – 313, 2008.

[62] P. Anbeek, K. L. Vincken, M. J. Osch, R. Bisschops, and V. Jeroen. "Probabilistic segmentation of white matter lesions in MR imaging". *NeuroImage*, 21(3):1037 – 1044, 2004.

[63] J. H. Morra, Z. Tu, A. W. Toga, and P. M. Thompson. "Automatic segmentation of MS Lesions using a contextual model for the MICCAI Grand Challenge". *The MIDAS Journal - MS Lesion Segmentation (MICCAI 2008 Workshop)*, 20(6), 2008.

[64] N. Shiee, P.-L. Bazin, A. Ozturk, D. S. Reich, P. A. Calabresi, and D. L. Pham. "A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions". *NeuroImage*, 49(2):1524 – 1535, 2010.

[65] D. L. Pham and J. L. Prince. "An adaptive fuzzy C-means algorithm for image segmentation in the presence of intensity inhomogeneities". *Pattern Recognition Letters*, 20(1):57 – 68, 1999.

[66] S. Z. Li. *Markov Random Field Modeling in Computer Vision*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1995.

[67] K. V. Leemput, F. Maes, D. Vandermeulen, A. Colchester, and P. Suetens. "Automated segmentation of Multiple Sclerosis lesions by model outlier detection.". *IEEE Transactions on Medical Imaging*, 20(8):677, 2001.

[68] R. Khayati, M. Vafadust, F. Towhidkhah, and M. Nabavi. "Fully automatic segmentation of multiple sclerosis lesions in brain MR FLAIR images using adaptive mixtures method and markov random field model". *Computers in Biology and Medicine*, 38(3):379 – 390, 2008.

[69] N. K. Focke, M. R. Symms, J. L. Burdett, and J. S. Duncan. "Voxel-based analysis of whole brain FLAIR at 3T detects focal cortical dysplasia". *Epilepsia*, 49(5):786–793, 2008.

[70] S. Knake, C. Triantafyllou, L. L. Wald, G. Wiggins, E. Halgren, and P. E. Grant. "3T phased array MRI improves the presurgical evaluationin focal epilepsies". *Neurology*, (65):1026–1031, 2005.

[71] A. M. Dale, B. Fischl, and M. I. Sereno. "Cortical surface-based analysis: I. Segmentation and surface reconstruction". *NeuroImage*, 9(2):179 – 194, 1999.

[72] R. S. Desikan, F. Sgonne, B. Fischl, and B. T. Quinn et al. "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest". *NeuroImage*, 31(3):968 – 980, 2006.

[73] B. Fischl, D. H. Salat, and E. Busa et al. "Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain". *Neuron*, 33(3):341 – 355, 2002.

[74] I. Aganj, G. Sapiro, N. Parikshak, S. K. Madsen, and P. M. Thompson. "Measurement of cortical thickness from MRI by minimum line integrals on soft-classified tissue". *Human Brain Mapping*, 30(10):3188 –3199, 2009.

[75] S. E. Jones, B. R. Buchbinder, and I. Aharon. "Three-dimensional mapping of cortical thickness using Laplaces equation". *Human Brain Mapping*, 11:12–32, 2000.

[76] S. Calinon. *Robot Programming by Demonstration: A Probabilistic Approach*. EPFL/CRC Press, Washington D.C., USA, 2009.

[77] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi. "A no-reference perceptual blur metric". *International Conference on Image Processing (ICIP)*, 3(5):357 – 360, 2002.

[78] C.-A. Yang, M. Kaveh, and B. Erickson. "Automated detection of Focal Cortical Dysplasia lesions using volume-based discriminative features". *Biomedical Imaging: Nano to Macro, 8th IEEE International Symposium on*, 2(5):865–870, 2011.

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

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

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

[82] S. Z. Li. *Markov random field modeling in image analysis*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2001.

[83] D. R. Chittajallu, S. K. Shah, and I. A. Kakadiaris. "A shape-driven MRF model for the segmentation of organs in medical images". *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, 2(6):3233 –3240, June 2010.

[84] J. Besag. "Spatial interaction and the statistical analysis of lattice systems". *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):192–236, 1974.

[85] Y. Zhang, M. Brady, and S. Smith. "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm". *Medical Imaging, IEEE Transactions on*, 20(1):45 –57, Jan. 2001.

[86] A. P. Dempster, N. M. Laird, and D. B. Rubin. "Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, series B*, 39(1):1–38, 1977.

[87] Y. Boykov and G. Funka-Lea. "Graph Cuts and Efficient N-D Image Segmentation. *International Journal of Computer Vision*, 70(2):109–131, Nov. 2006.

[88] Y. Boykov and M. P. Jolly. "Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images". *Proceedings Eighth IEEE International Conference on Computer Vision ICCV 2001*, 1:105–112, 2001.

[89] J. M. Chambers. *Graphical methods for data analysis*. Chapman & Hall statistics series. Wadsworth International Group, New York, New York, USA, 1983.

[90] E. L. Lehmann and J. P. Romano. *Testing Statistical Hypotheses*. Springer Texts in Statistics. New York, New York, USA, 2005.

[91] W. J. Krzanowski and D. J. Hand. *ROC Curves for Continuous Data*. Monographs on statistics and applied probability. Washington D.C., USA, 2009.

[92] R. Wang and K. Tang. Feature selection for maximizing the area under the roc curve. *Data Mining Workshops, IEEE International Conference on*, 1(1):400 –405, Dec. 2009.

[93] J. J. Bartko. Measurement and reliability: statistical thinking considerations. *Schizophrenia Bull*, 17:483489, 1991.