Identifying Lesions in Paediatric Epilepsy using Morphometric and Textural analysis of Magnetic Resonance Images

by

Sancgeetha Kulaseharan

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

Master of Science

in

The Faculty of Science

Modelling and Computational Science

University of Ontario Institute of Technology

December 2017

© Sancgeetha Kulaseharan, 2017

Abstract

We develop an image processing pipeline on Magnetic Resonance Imaging (MRI) sequences to identify features of Focal Cortical Dysplasia (FCD) in patients with MRI-visible FCD. We aim to use a computer-aided diagnosis system to identify epileptogenic lesions with a combination of established morphometric features and textural analysis using Gray-Level Co-occurrence Matrices (GLCM) on MRI sequences. The model will be validated on paediatric subjects. Preliminary morphometric analysis explored the use of computational models of established MRI features of FCD in aiding identification of subtle FCD on MRI-positive subjects. Following this, classification techniques were considered. The 2-Step Naive Bayes classifier was found to produce 100% subjectwise specificity and 94% subjectwise sensitivity (with 75% lesional specificity, 63% lesional sensitivity). Thus it correctly rejected 13/13 healthy subjects and colocalized lesions in 29/31 of the FCD cases with MRI visible lesions, with 63% coverage of the complete extent of the lesion using supplied lesional labels.

Acknowledgements

First, I would like to thank my supervisors Dr. Mehran Ebrahimi and Dr. Elysa Widjaja for their guidance and support throughout this project and The Hospital for Sick Children (SickKids), Toronto, Canada for providing the data. I would also like to thank the members and faculty of the Modelling and Computational Science program, particularly my fellow members of the Imaging Lab. Finally, I would like to thank my family and friends for their support and encouragement.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Focal Cortical Dysplasia

With an incidence rate of 4-9/1000 per year, epilepsy is characterized as one of the most frequently observed neurological disorders in childhood [CNP+06] [CJGM00]. Of these cases, 25% are drug resistant and identified as cases of intractable epilepsy. Prolonged uncontrolled epilepsy is detrimental to the neurodevelopment of a child and can cause long-lasting impact on their neurocognitive development or premature death. Surgical treatment through resection of the lesion responsible for epilepsy is an option in selected patients. Pre-surgical evaluation of Magnetic Resonance Imaging (MRI) plays a crucial role in identifying the lesion.

Focal Cortical Dysplasia (FCD) is a brain malformation that is frequently responsible for epilepsy in children and accounts for approximately 26% of intractable epilepsy in children [WCK+98]. In children with refractory focal epilepsy, a lesion can be identified on MRI in 30-85% of cases [Dun97] [CTG+03]. FCD is a brain malformation that causes seizures [RL01]. Unlike tumors, FCD can be subtle and difficult to detect using standard MRI. FCD lesions are associated with several T1-weighted MRI features, including abnormal cortical thickness, abnormal sulcal pattern, blurring of the gray-white

matter junction, and abnormal signal in the white and/or gray matter (Figure 1.1). A radiologist may employ these features to identify an FCD on MRI [ACB+03].

The use of different modalities is another consideration in lesion detection. Different MRI acquisitions methods such as volumetric T1-weighted, T2-weighted, and FLAIR (Fluid Attenuated Inversion Recovery) vary in terms of their ability to recognize certain features. For instance, studies have indicated T2-weighted, and FLAIR MRI images are better at identifying abnormal signal in the white matter compared to sequences acquired from T1 [FSBD08]. On the other hand, volumetric T1-weighted images, due to their thinner slices, are superior for examining variations in cortical thickness [DCDDS12]. FCD is one of the most challenging group of lesions to detect on conventional MRI as the imaging features may be subtle and not infrequently missed. Additionally, the commonly recognized features of FCD may not all be present in a given case. Some cases of FCD may exhibit one or two of the identifiable features, while others may exhibit more [ACB+03]. For these reasons, limiting assessment to one feature could inevitably lead to false negative. Those with FCD and normal MRI are known to have MRI-negative epilepsy, or non-lesional epilepsy. Cases with visible lesions on MRI have MRI-positive epilepsy [ACB+03]. Up to 72% of cases of MRI negative epilepsy have FCD [LLK+05] [JNB+07] [MBR+07] [KHD+07].

Patients with a lesion seen on MRI have better surgical outcomes, while those with MRI-negative focal epilepsy have poorer surgical outcomes. In studies of MRI-negative cases of focal epilepsy, the surgical resection seizure cure rates fell within the range of 0-31% [HW88] [KCB+03] [Ras63]. Comparing this to a study of 294 subjects of which the majority was composed of MRI-positive cases (91% of the total subjects) of lesional epilepsy, there was a 73% seizure cure rate following neocortical resection [Bon80]. Thus, pre-surgical identification of lesion can benefit the surgical outcome and has therapeutic and prognostic implications [AWG+17], [ACB+03].

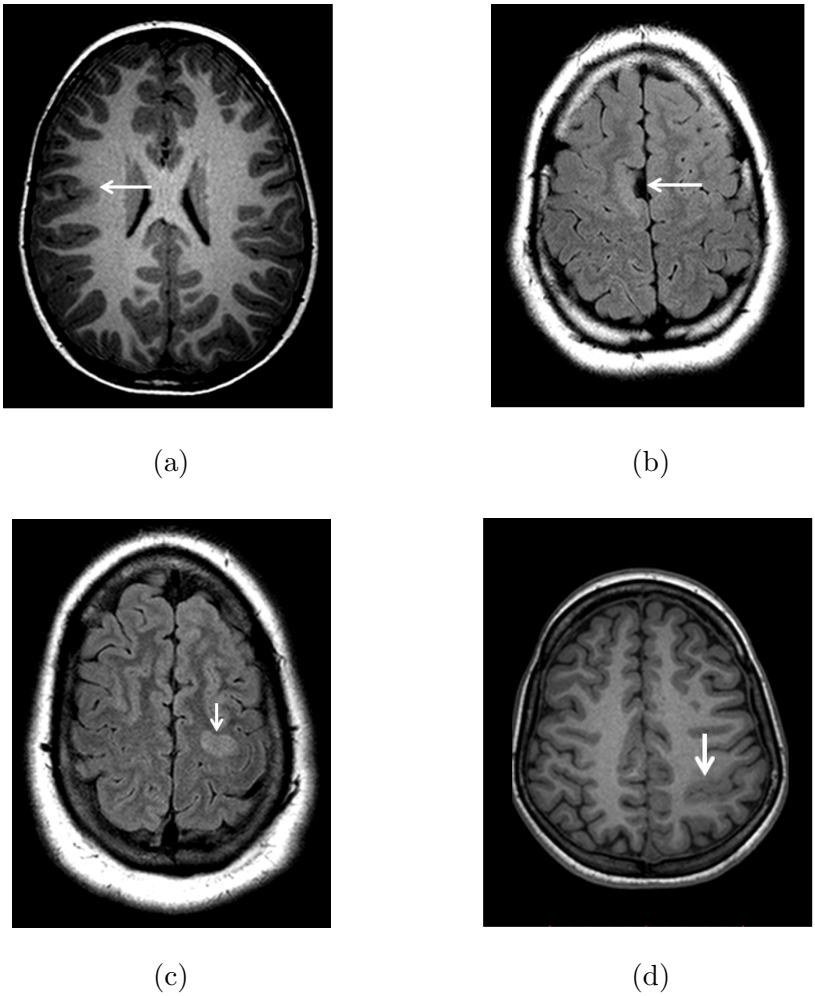


Figure 1.1: MRI features of FCD used in standard neuroradiological analysis. Recognized FCD features include blurring of the gray-white matter junction, abnormal signal in the white or gray matter, abnormal sulcal pattern, and cortical thickening. Sample cases with arrows indicating regions of abnormality on the basis of standard MRI features used for diagnostic analysis. (a) Blurring Gray White Junction and also high T1 signal in cortex (Subject NL005), (b) Abnormal Sulcation (Subject NL036), (c) High signal in cortex (FLAIR) (Subject NL037) and (d) Abnormal Cortical Thickness (Subject NL039). Two cases of high signal are presented in (a) and (c). The reason for doing so is to depict that even in cases where the same feature is present, the magnitude at which it presents itself may vary from subject to subject.

Differentiating FCD lesions from healthy cortex using standard radiological analysis does not always yield results and serves as motivation for quantitative image analysis [LWN08]. The subjectivity of visual assessment and reliance on observer expertise, highlights the need for an objective tool for MRI assessment. The availability of higher resolution MRI acquisitions have resulted in the natural step towards the use of computer-aided techniques during pre-surgical evaluation. Incorporating image-processing techniques offers the potential to detect subtle structural variations not visible otherwise, and improve prognosis [ACB+03]. Steps to produce automated lesion detection through post-processing techniques using multiple features in conjunction with classification using surface based features [AWG+17] [HKS+14] and textural analysis [ACB+03] have yielded positive results. Such methods have been shown to improve pre-surgical detection of Focal Cortical Dysplasia in MRI scans [ACB+03].

1.2 Literature Review

Many efforts have been made to improve epileptogenic lesion detection through computational means. [EAHCL13] explored the use of a one-class Support Vector Machines (OC-SVM) in voxel-wise classification for intractable epilepsy. Their dataset consisted of 29 Healthy control subjects (18-53 years), and the test set consisted of two FCD patients. Following normalization, Statistical Parametric Mapping (SPM) was used to label tissue classes. Z-score values were computed from probability maps of the control data for feature data. Given that SVM normally follows a standard two class problem, a modified version [SPST+01] was employed in which samples are trained solely on controls before being mapped to a higher dimensional space using a feature map. This yielded distance maps, which is finally thresholded at appropriate values to identify lesions. Voxel clusters were returned as output and ranked on the basis of size and suspicion.

[FSBD08] also utilized SPM, in this case, to perform voxel-based analysis of T2 FLAIR

scans that were coregistered (i.e. aligned to T1-weighted acquisitions) for a dataset of 25 FCD cases and 25 controls subjects to automate detection. Scans were then segmented into tissue classes and normalized. To evaluate for differences, regions of interest were manually outlined in control cases for five anatomical areas and these regions were compared to their counterparts in FLAIR images. SPM was used in analysis to compare individual controls against other control subjects and to compare individual patients against control data.

Whole-brain morphometry was explored for a dataset of 41 patients specifically with frontal lobe epilepsy with 41 matched controls by [HBS+16]. Structural differences were mapped relative to healthy subjects, and comparison was performed based on variations in cortical thickness and folding complexity. Classification was accomplished using a SVM classifier.

To enhance visualization of MRI features of FCD, specifically blurring at the gray-white interface, [HGF+05] developed an automated morphometric analysis script in MATLAB. Subsequent to normalization and segmentation, voxels with signal intensity between gray and white matter are identified with a histogram analysis of the segmented gray matter and white matter images. Binary images are composed to represent the distribution of the voxels. These images are filtered by performing a three-dimensional convolution and before finally being compared to the normal controls.

Exploring Surface Based Morphometry (SBM), through tools like FreeSurfer can achieve 3D-Cortical Mapping and aid in automated detection of FCD lesions [MBF15]. Textural analysis can also supply information not directly available. [BAC+01] applied a post-processing method combining morphometry and texture that supplies quantitative information referencing spatial gray level variations in pixel neighborhoods to produce maps highlighting cortical thickening, blurring at the junction and relative variations in signal intensity for cases of extra-temporal partial epilepsy. [BBC+08] too combined attributes of morphometry and texture with a fuzzy K-Nearest Neighbor (fk-NN) clas-

sifier. The combination of textural analysis and structural features was also explored by [ACB+03] who combined the two with a 2-Step Bayesian classifier. The process consisted of creating 3D maps of cortical thickness, gradient magnitude at the interface, and a relative intensity operator along with textural analysis using Gray Level Co-Occurrence Matrices (GLCMs). Classification was conducted in two steps using a Naive-Bayes classifier and results were thresholded using control data to identify lesional clusters.

Computer-assisted implementations of automated Focal Cortical Dysplasia epileptogenic lesion detection have shown positive results, however these methods have not produced equal success when applied to cases in developing brains [AWG+17].

1.3 Objective

FCD is frequently responsible for surgically remediable cases of intractable epilepsy in children, however lesion detection requires a radiologist to examine a scan slice by slice, and the lesion itself is very subtle. In studies conducted with adult cohorts, FCD detection was improved with computational techniques, but they have not been as effective when applied to developing brains [AWG+17]. Thus, our key aim is to assist radiologists in paediatric lesion identification in MRI-positive cases using computational techniques. We seek to explore the development of a unified mathematical model that considers segmentation, morphometric analysis, and textural analysis in a joint formulation for this purpose. The model will be validated on paediatric subjects. This study seeks to develop an image processing pipeline using textural and morphometric analysis on MRI sequences to identify features of FCD in patients with MRI-visible FCD. We will examine statistical, machine learning, and image processing techniques as means of enhancing diagnostic practices for paediatric Focal Cortical Dysplasia.

We seek to develop a system that could determine whether a subject is healthy or has FCD, and detect the location of FCD in a subject identified as FCD-positive. We aim

to examine the use of objective features including cortical thickness, and blurring of the interface, with textural features to identify features of FCD in patients with MRI-positive FCD. Regions will be identified through a combination of thresholding and classification algorithms. Following this, a supplementary experiment will be performed extending classification techniques to MRI-negative epilepsy to determine if the tool can detect subtle features FCD not detected by visual assessment.

1.4 Outline

The work conducted in [ACB+03] served as the key motivation of our analysis upon which a majority of our pipeline was built. Our goal is to consider more than one morphometric feature with textural attributes to examine their use in diagnostic methods. The background information pertaining to software used, a description of the data set, the measures used for evaluating performance, an introduction to the common T1 MRI features, and the theory relating to the classification methods and textural analysis are first presented in Chapter 2. Following this, Chapter 3 contains the experiments and results. The experiments can be split into two major ideas: preliminary morphometry experiments (Section 3.1) and identifying epileptogenic lesions using classification techniques using morphometric and textural analysis (Section 3.2).

In Section 3.1, we begin with preliminary data analysis using morphometric features. Specifically, we seek to examine the use of cortical thickness measures (Subsection 3.1.1) and blurring of the gray/white matter interface using gradient magnitude (Subsection 3.1.2) on T1 acquisitions as a means of identifying epileptogenic features in MRI-Positive subjects. A secondary experiment was performed using blurring of the gray/white matter to compare results produced from T2-weighted and FLAIR images.

Voxelwise classification has been a key method for detecting lesions in medical imaging. Building off this idea we seek to incorporate textural analysis. Thus our next step

is to consider morphometry and texture in analysis conducted on MRI-Positive subjects and Healthy Controls. To perform classification we seek to produce a modified method of the 2-Step Bayesian classifier described by [ACB+03]. In addition to this, we perform comparative analysis using other common classification techniques. In 3.2 we start with the background 3.2.1 and experimental setup 3.2.2 for our experiments in implementing classification techniques. A preliminary comparative study 3.2.3 of several classification techniques is performed to act as an assessment evaluator for subsequent implementations. 2-step methods using two different implementations (2-Step Naive Bayes classifier, and a 2-Step SVM classifier) are found in 3.2.4 and 3.2.5. Finally techniques are evaluated for performance 3.2.6, with the optimal method being applied in a supplementary case study of MRI negative cases 3.2.7 and an additional experiment using this technique to conduct comparison for classification using T1, T2 and FLAIR acquisitions 3.2.8. Chapter 4 details the conclusion and discusses possible future work.

Chapter 2

Background Information

2.1 Software

Research was conducted on an Ubuntu operating system using R, MATLAB in combination with SPM (Statistical Parametric Mapping), and FreeSurfer software packages.

FreeSurfer Neuroimaging analysis software is freely available and developed specifically for processing brain MR images and quantifying Surface Based Morphometry (SBM). FreeSurfer v5.3.0 ([DFS99][DS93] [FD00] [FLD01] [FSB$^+$02]; [FSvdK$^+$04][FSD99] [FVDKD$^+$04] [HJS$^+$06] [JCG$^+$06]; [SDB$^+$04] [RRF10][RSRF12]) was employed for preprocessing T1-weighted images for all cases and controls in order to perform cortical reconstruction, produce anatomical labelling and structure details, and to co-register T2 and FLAIR scans to T1-weighted images. The OpenMP flag was used to run subject processing on 4 cores, resulting in a runtime of approximately 6-9 hours per subject. As a brief overview of the FreeSurfer pipeline, given a T1 weighted image, the native resolution is interpolated to 1mm^3 isotropic voxels, this is followed by intensity normalization, skull stripping using a watershed algorithm [SDB$^+$04], volumetric labeling, white matter segmentation, surface extraction, surface atlas registration, and surface labeling [Per07]. FreeSurfer uses .mgz, a file type unique to the software that is similar to nifti (a stan-

dard file format used to save volumetric MRI data), to store compressed high-resolution structural data and other information that may be overlaid on the structural volume. The annotation files are output with the .annot extension. FreeSurfer supplied scripts were used for loading these volumes into MATLAB.

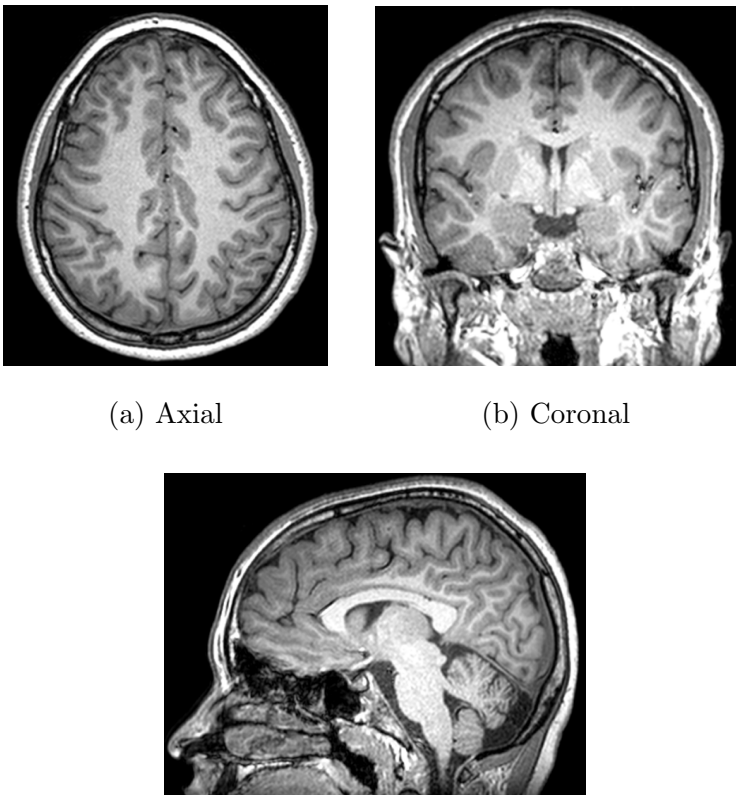
MATLAB was selected for implementation of preliminary experiments, and pre- and post- classification data processing due to its built-in imaging capabilities, as well as the bidirectional capacity to read and write FreeSurfer generated volume and surface data. The SPM (Statistical Parametric Mapping) software is a suite of MATLAB developed for neuroimaging that was additionally employed for the processing of FLAIR and T2 images.

Due to the size of the data considered in experiments, performing classification in MATLAB was not a viable option. Thus, R (v3.4.1) was used to implement classification techniques. R is an open-source system of programming for statistical computation and data mining applications. The program can handle large-scale data, and the CRAN (Comprehensive R Archive Network) repository offers a large library of additional modules, including numerous predictive modelling packages, that can be used in conjunction with R.

2.2 Data Description

Data is available as Digital Imaging and Communications in Medicine (DICOM), a standard format in medical imaging. The volumetric T1-weighted, T2-weighted, and FLAIR imaging was acquired on 3 Tesla (3T) MRI in 32 children with focal epilepsy having subtle lesion that was suspected to be FCD (MRI positive). The mean age of the patients was 11.73 years. Subjects were supplied labelled with a format NL0**. Subject NL037 was omitted due to error at the acquisition stage. An additional 4 cases of MRI-negative subjects were also included for a secondary experiment. All patients underwent

epilepsy surgery resection and had post-operative Computed tomography (CT) or volumetric T1 MRI. Healthy control data was acquired for 13 children for which T1-weighted, T2-weighted, and FLAIR images were available. Sample images are displayed in Figure 2.1 and Figure 2.2.



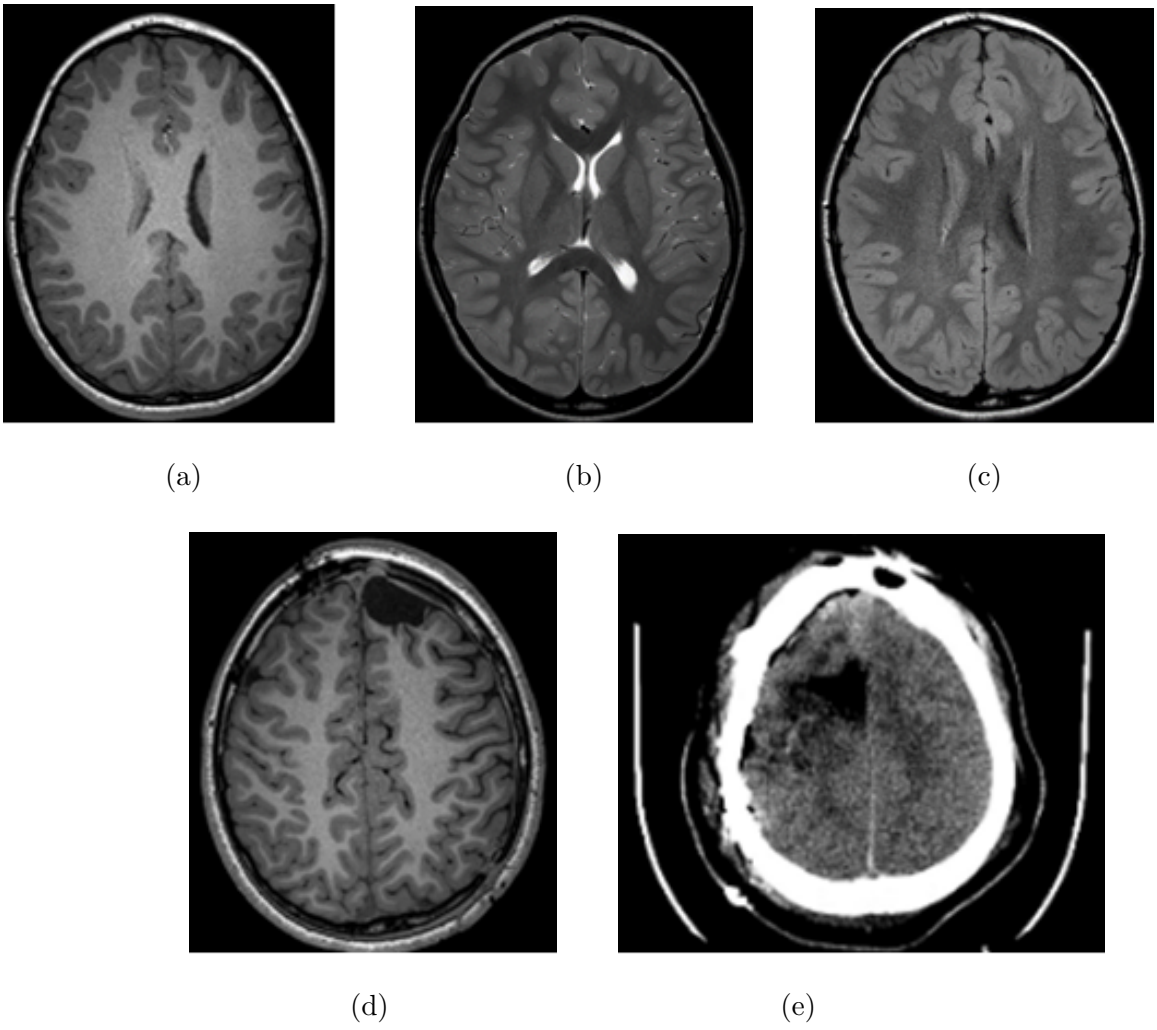
(a) Axial (b) Coronal

(c) Sagittal

Figure 2.1: Sample T1-weighted acquisition. (a) displays the axial slice, (b) the coronal slice and (c) the sagittal slice of the T1 acquisition of subject NL004.

The MRI-positive patient case data including ILAE (International League Against Epilepsy) surgical outcome, histology information, and suspected lesion location using FreeSurfer labels relating to pre-surgery neuroradiological work-up reports can be found in Table 2.2. ILAE Surgical outcome refers to a scale from Class I to V (Class I: Completely seizure free; Class II: Only auras, no other seizures; Class III: 1-3 seizure

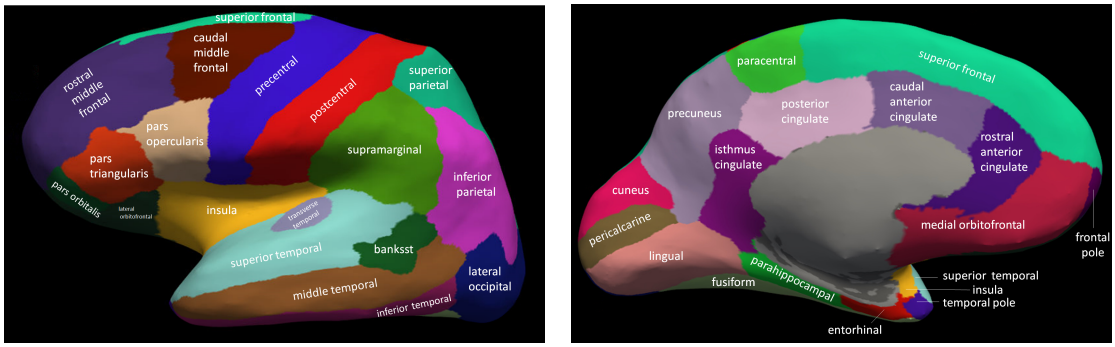
days/year; Class IV: 4 seizure days/year - 50% reduction in baseline number of seizure days; Class V: <50% reduction in baseline number of seizure days - 100% increase in baseline number of seizure days) detailing patient improvement following surgical resection [WBF+01]. Details are obtained from histology reports made by pathologists examining resected brain tissue post-surgery under a microscope to identify the type of lesion. MRI-negative details are found in Table 2.3. The location of lesion in MRI-negative patients was identified using post-surgical resection imaging. The brain can be coarsely labelled using 8 possible locations (4 per hemisphere): right and left frontal, temporal, parietal, and occipital. Fine labelling corresponds to the brain being labelled to 70 possible locations, 35 for each hemisphere (refer to Figure 2.3 and Table 2.1). The histology reports supplied labels were corresponded to anatomical labels derived using FreeSurfer. Note that the listed label titles in Table 2.1 correspond to the FreeSurfer default naming and the insula falls under both the frontal and temporal lobes.



(a) (b) (c)

(d) (e)

Figure 2.2: Pre-surgery and post-surgery supplied modalities. Row 1 displays pre-surgery supplied imaging sequences: (a) T1-weighted, (b) T2-weighted, and (c) FLAIR. Row 2 presents post-surgery supplied modalities with resected regions visible: (d) T1-weighted and (e) CT.



(a) (b)

Figure 2.3: Left hemisphere fine labelling using FreeSurfer labels.

Frontal	Parietal	Temporal	Occipital
superiorfrontal	precuneus	inferiortemporal	rightcuneus
rostralmiddlefrontal	inferiorparietal	bankssts	rightlateraloccipital
caudalanteriorcingulate	isthmuscingulate	entorhinal	rightlingual
caudalmiddlefrontal	postcentral	fusiform	rightpericalcarine
lateralorbitofrontal	posteriorcingulate	middletemporal	
medialorbitofrontal	superiorparietal	parahippocampal	
paracentral	supramarginal	superiortemporal	
parsopercularis		temporalpole	
parsorbitalis		transversetemporal	
parstriangularis		insula	
precentral			
rostralanteriorcingulate			
frontalpole			
insula			

Table 2.1: Coarse and fine labelling

Case label	ILAE Surgical Outcome	Histology	Location of lesion R-right; L-left
NL001	II	Non-specific gliosis	L Superior temporal, Middle temporal, Temporal pole
NL002	I	FCD type IIA	R Inferior temporal, Fusiform
NL003	I	FCD type IIA	L Superior frontal
NL004	I	FCD type IIB	L Inferior frontal, Fusiform
NL005	I	Non-specific gliosis	R Precentral
NL007	I	FCD type IIB	L Insula
NL008	I	non-specific gliosis	L Parsopercularis
NL009	III	Oligodendrogliosis	L Temporal pole, Superior temporal, Middle temporal
NL010	II	Non-specific gliosis	L Superior frontal
NL011	I	Oligodendrogliosis	L Parahippocampal, Fusiform
NL012	I	Oligodendrogliosis	R Lateral orbital frontal, Medial orbital frontal
NL013	I	FCD type IIB	R Inferior parietal, Lateral occipital
NL016	III	Non-specific gliosis	L Temporal pole, Superior temporal, Middle temporal
NL017	I	Oligodendrogliosis	L Temporal pole, Superior temporal, Middle temporal
NL018	I	Oligodendrogliosis	L Superior frontal, Rostral middle frontal, Caudal middle frontal, Parsopercularis, Parsorbitalis, Parstriangularis, Lateral Orbitofrontal

NL019	I	Non-specific gliosis	L Temporal pole, Superior temporal, Middle temporal
NL020	I	FCD type IIB	R Precentral, Rostral Middle frontal, Caudal Middle frontal
NL021	I	FCD type IIA	L Superior frontal
NL022	III	FCD type IIA	R Superior frontal, Rostral middle frontal, Caudal middle frontal, Parsopercularis, Parstriangularis, Lateral Orbitofrontal
NL023	I	FCD type IIB	L Postcentral
NL024	I	Non-specific gliosis	L Paracentral, Posterior cingulate
NL026	I	Oligodendrogliosis	R Superior frontal, Rostral middle frontal, Caudal middle frontal, Parsopercularis, Parstriangularis, Lateral Orbitofrontal
NL027	III	FCD type I	R Transverse temporal, Superior temporal, Middle temporal, Inferior temporal, Inferior parietal, Lateral occipital
NL029	I	FCD type IA	L Transverse temporal, Superior temporal, Middle temporal, Inferior temporal, Fusiform, Parrahippocampal gyrus, Lingual Pericalcarine, Cuneus, Lateral occipital
NL031	V	FCD type IB	L Temporal pole, Superior temporal, Middle temporal

NL032	IV	Oligodendrogliosis	L Temporal pole, Superior temporal, Middle temporal
NL033	I	FCD type I	L Superior frontal, Rostral Middle frontal, Caudal Middle frontal, Parsopercularis, Parstriangularis, Lateral Orbitofrontal
NL035	V	FCD type I	R Precuneus
NL036	I	FCD type IIA	R Superior frontal, Paracentral
NL037	IV	FCD type IIB	L Precentral, Superior frontal
NL038	I	FCD type IIA	L Temporal pole, Superior temporal, Middle temporal, Inferior temporal, Transverse temporal
NL039	I	FCD type IIB	L Precentral, Postcentral, Superior parietal, Supramarginal

Table 2.2: MRI-positive data

Case label	Location of lesion R-right; L-left
NL041	R Precentral, Postcentral, Parsopercularis, Parstriangularis, Middle temporal, Inferior temporal
NL042	L Postcentral
NL043	R Superior frontal
NL044	L Middle temporal, Inferior temporal, Postcentral, Superior parietal

Table 2.3: MRI-negative data

2.3 Methods of Measuring Performance

Standard classification performance evaluation utilizes sensitivity (Eq. 2.1) and specificity (Eq. 2.2) measures. Sensitivity measures the true positive probability and specificity the true negative probability.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2.1}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{2.2}$$

where TP stands for the true positive count, FN stands for the false negative count, TN stands for the true negative count, and FP stands for the false positive count. The confusion matrix is presented in Table 2.4.

		PREDICTION	
		Positive	Negative
ACTUAL	Positive	TP	FP
	Negative	FN	TN

Table 2.4: Confusion matrix. Visualization of classification performance using a confusion matrix to depict relationship between predicted results and actual values, given class labels Positive and Negative.

Thus we require some standard for evaluating what the actual values are. Lesion locations corresponding to true regions were derived from pre-surgical radiological reports. Sample cases for FCD patients NL008 and NL029 are presented respectively in Figure 2.4 and Figure 2.5. With regards to classification, if the structures selected either partially or wholly colocalize the true lesion in a patient with FCD, we consider the lesion to be identified.

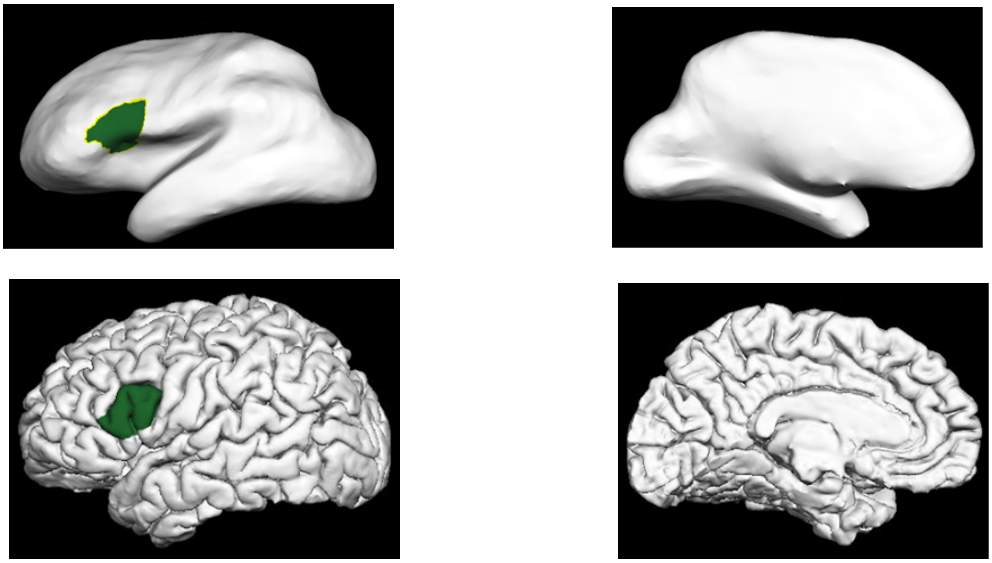


Figure 2.4: Lesion Labelling Patient NL008 on inflated and pial surfaces

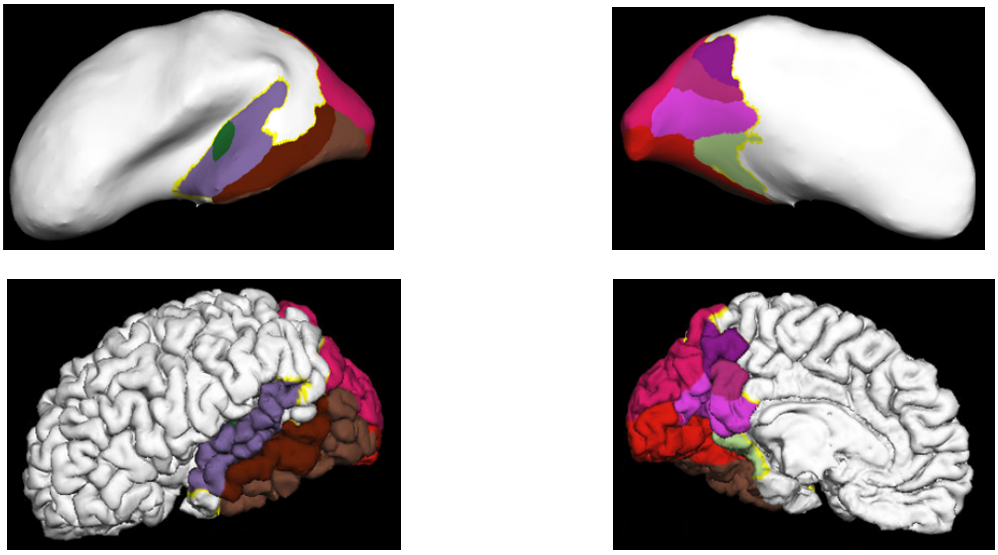


Figure 2.5: Lesion Labelling Patient NL029 on inflated and pial surfaces

Subjectwise sensitivity (Eq. 2.3) is the sum of all FCD subjects in which a lesion is correctly identified divided by the total number of FCD subjects. Subjectwise specificity (Eq. 2.4) is the sum of all control subjects in which no lesion is identified divided by the total number of control subjects. A structure is considered lesional if it was specified

in the pre-surgical radiological report for the subject. Lesional sensitivity (Eq. 2.5) is measured as the sum of all structures labeled lesional by the classifier for all subjects divided by the total structures labelled lesional for all subjects in pre-surgical reports. Lesional specificity (Eq. 2.6) is the sum of all structures labeled non-lesional by the classifier for FCD subjects divided by the total structures labelled non-lesional for FCD subjects in pre-surgical reports. We will consider subjectwise sensitivity and subjectwise specificity as our key performance measures, and use the lesional sensitivity and lesional specificity as supplementary information for analysis [ACB+03]. Mathematically defining these measures, for N_F number of FCD patients and N_C number of healthy control subjects,

$$\text{Subjectwise Sensitivity} = \frac{\sum_{i=1}^{N_F}[(TP)_i > 0]}{N_F} \tag{2.3}$$

$$\text{Subjectwise Specificity} = \frac{\sum_{i=1}^{N_C}[(TN)_i > 0]}{N_C} \tag{2.4}$$

$$\text{Lesional Sensitivity} = \frac{\sum_{i=1}^{N_F}(TP)_i}{\sum_{i=1}^{N_F}[(TP)_i + (FN)_i]} \tag{2.5}$$

$$\text{Lesional Specificity} = \frac{\sum_{i=1}^{N_C}(TN)_i}{\sum_{i=1}^{N_C}[(TN)_i + (FP)_i]}. \tag{2.6}$$

In these equations $(TP)_i$ is the true positive relating to the i-th subject. The other quantities are similarly defined. We are predominantly interested in subjectwise performance, that is to say, a technique's ability to not select any lesional structures in control cases and partially/wholly colocalize lesions in FCD patients. However the extent of the lesion can differ vastly from case to case, as depicted in Figures 2.4 and 2.5. For this reason, we incorporate lesional measures as additional measures to overcome the subjectwise measures limitations in localization.

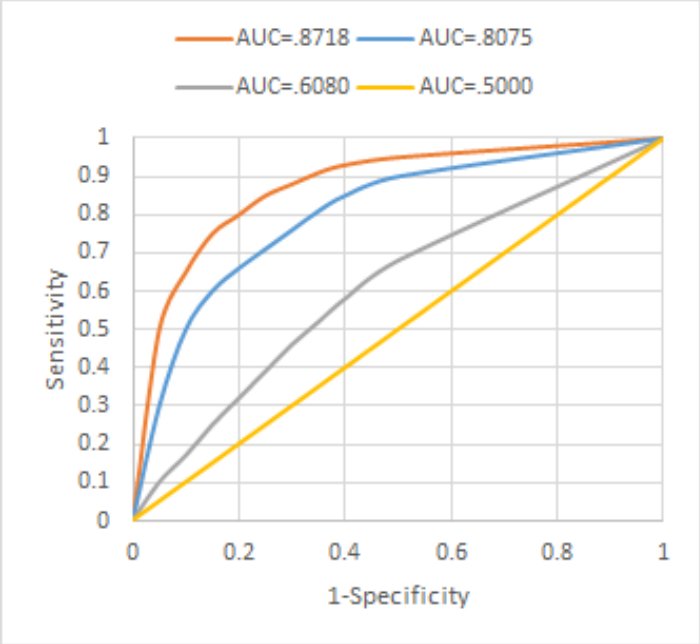


Figure 2.6: ROC curves and AUC. A ROC curve examines the trade-off between clinical sensitivity and specificity for each possible threshold cut-off. Computing the AUC evaluates the accuracy of a test. Sample AUC curves are presented. The AUC may fall between 0.5 and 1, with satisfactory values falling between 0.7 and 1. The closer the AUC value is to 1, the more accurate the test.

Parameter selection plays a vital role in many image processing techniques and impacts the sensitivity and specificity. Thus, selecting optimal parameters to improve performance is an important consideration (Figure 2.7). A receiver operating characteristic (ROC) curve is a standard tool in medical decision making and can be used to improve predictive performance through criterion selection. The ROC curve plots the false positive fraction (expressed by 100%-specificity) against sensitivity for different values of a parameter and allows the user to select the optimal decision threshold. Each point on an ROC is a representation of a selected cut-off threshold. The Area Under the ROC Curve (AUC) is used to measure the accuracy of the test. An AUC=1 implies perfect separation, and AUC=0.5 coincides to a diagonal line and implies the test is worthless.

A satisfactory AUC falls within the range of 0.70 to 1. Thus the closer the curve is to the upper left corner, the better the performance (Figure 2.6). ROC curves offers a graphical representation of the trade-off between clinical sensitivity and specificity and the AUC indicates if the test itself is beneficial (discriminative power) [Eke12]. The AUC can be computed using the trapezoid method such that

$$\text{AUC} = \sum_{k=1}^{n} (X_{k+1} - X_k) \frac{(Y_k + Y_{k+1})}{2} \tag{2.7}$$

where for the function $f(X)$ and $n = 1, ..., k$ segments, $(X_{k+1} - X_k)$ denotes the spacing between a consecutive pairs of points with $Y_k = f(X_k)$ and $Y_{k+1} = f(X_{k+1})$.

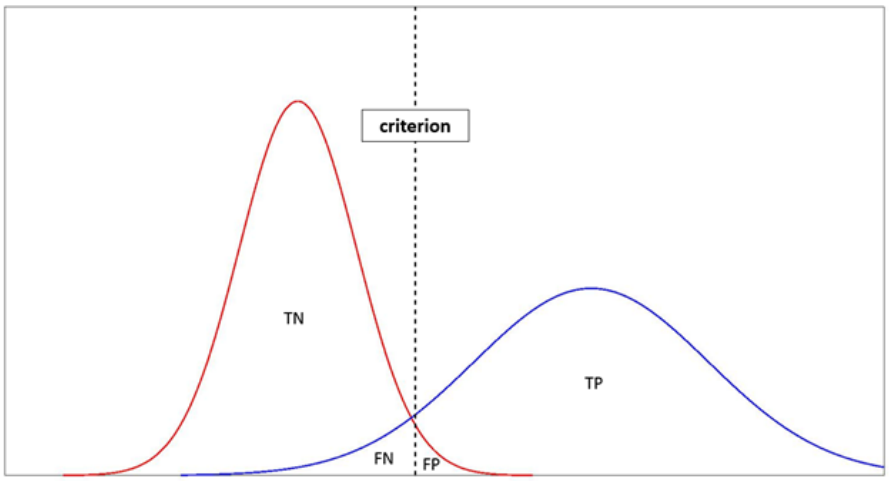


Figure 2.7: Criterion selection. Consider a binary classification problem with class labels lesional and non-lesional. The diagnostic performance is measured by the ability to distinguish these two groups from one another. The value of the criterion acts a decision threshold. Cases to the right of the line are identified as lesional, and cases to the left non-lesional.

Youden (1950) presented the Youden index (Eq. 2.8) as a means of selecting the optimal criterion for trade-off. J can be expressed graphically as the maximum vertical distance between a diagonal line and the ROC curve. The associated parameter for the

J is selected as optimal

$$J = \max\{\text{sensitivity}_c + \text{specificity}_c - 1\} \tag{2.8}$$

for possible criterion values c, with the aim to maximize the J value [You50][Eke12].

2.4 Common Features of T1-weighted MRI – Cortical Thickness and Blurring of Gray-White Matter Interface

High-resolution medical imaging depict considerable detail and enables in-vivo morphometric analysis. Cortical thickness and blurring at the gray-white matter boundary are two of the established features of Focal Cortical Dysplasia lesions on volumetric T1-weighted images [HDVA+08]. Additionally, they are advantageous due to the fact that models derived from these features operate in three dimensions. The implication being that information can be considered simultaneously from consecutive slices and multidirectionally. This is a benefit over standard radiological analysis which may use slice-by-slice visualization [ACB+03].

Cortical thickness serves as a useful measurement in offering insight in regards to anatomical abnormality [HDVA+08] [ACB+03]. The cortical thickness in a normal brain can range from 2 mm to 4 mm, with the average falling somewhere between 2.5 and 3 mm. Standard variations in cortical thickness may be attributed to several factors including location in functionally unique regions, or differences in cell types [HDVA+08].

In cases with lesions, the transition from gray- to white- matter loses distinction and may be less visible. For this reason, blurring at the interface between the gray and white matter is a common feature characterized by a "fogginess" on volumetric T1-weighted MRI. Enhancing visualization of this boundary using modern post-processing can flag regions lacking distinction and has produced improved detection rate [HGF+05].

2.5 Classification Techniques

2.5.1 Introduction to Classification

In classification, given a set of input variables, that influence one or more outputs, the aim is to predict these outputs. We will refer to these inputs as features, and the outputs as class labels.

Take $\mathbf{X}$ to be a n-dimensional vector containing feature data such that $\mathbf{X} = (X_1, ..., X_N)$. X_n refers to the n-th feature. Let $\mathbf{C} = (C_1, ..., C_M)$ be a vector of class labels. C_m refers to the m-th class label. If $M = 2$ it is a binary classification problem. We seek to use a classifier to map the feature data to the class data. In unsupervised learning, initial known labels are not available and inferences are made solely off input to express patterns and associations amongst the feature measures. We will be considering supervised classification techniques, where predictions of new feature vectors are made on the basis of known pairings of feature vectors and class labels.

A training set is employed to build the model, and a testing set is used for validation. The training set contains N paired sets of features and response classes, and the testing set contains feature samples alone with class labels to be determined. Splitting data into model training and validation sets ensures that prior knowledge is not directly passed to predictions. This is referred to as the hold-out method and is the simplest cross-validation implementation. In cross-validation, some of the data is omitted from the training set and instead used for testing purposes. In an ideal situation the training and testing samples would be two unique sets, however when sufficient data is not available to do so, other techniques are employed. In K-fold cross-validation the data is split into K approximately equal portions, and for each data portion, the data is fit to the other $K - 1$ sets. This process is repeated for each $k = 1, 2, 3, \ldots, K$. When $K = N$ the process is referred to as leave-one-out cross-validation. Leave-one-out cross-validation is higly unbiased, however can be computationally expensive and have high

variance [HTF09] [Mur12].

In medical imaging, classification is an especially useful tool for objects that may not be simply represented, as is the case with FCD lesions [Suz12]. Computational models to improve visual lesion detection combining recognized features of FCD with complimentary textural features have produced positive results [ACB+03] [AWG+17]. Directly using voxel information, as is done when computing textural features, can circumvent errors that may arise when using values are computed from segmented regions and has the potential to outperform other methods [Suz12] [HS+73]. When classifying images, texture is an especially significant distinguishing attribute to consider. We seek to employ classification techniques to combine the morphometric and textural information to select lesional locations and compare performance of classifiers. We will consider K-Nearest Neighbours, conditional inference trees, random forests, support vector machines and Naive-Bayes Classifiers.

2.5.2 k-Nearest Neighbour (k-NN)

The general assumption of k-Nearest-Neighbours (k-NN) is that similar inputs will result in similar outputs, and we can classify data by utilizing a measurement of this similarity. A non-parametric method, k-NN does not fit data to a model, but is instead memory-based. It is not necessarily a machine learning algorithm but a pattern recognition method. To employ k-NN for classification, given training data with known labels and a new unlabeled test input $\mathbf{X}$, a similarity measure is used to identify the k nearest points $\mathbf{X}_1, ..., \mathbf{X}_k$ in the training set. Among these k entries, the most common class label is selected to be the predicted label for the new data instance, Figure 2.8. In the event a tie for the majority vote occurs, it is broken at random.

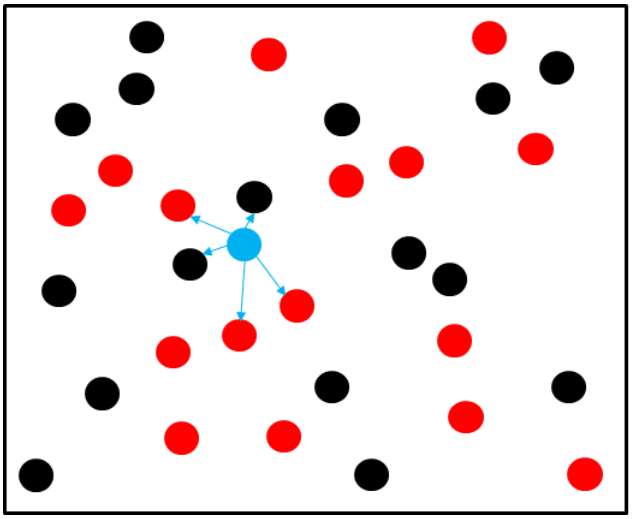


Figure 2.8: k-NN example for $k = 5$. Here were have a training set and two class labels visually presented as black and red. For a new unlabelled data (represented in blue) we seek the 5 nearest neighbours. The class is selected using a majority vote, in this case leading to a label of red.

k-NN is easy to implement, and works well in cases where the decision boundary between data classes is irregular. A caveat of the method is that it requires retention of the complete training set. Variations can be made to parameter k and the similarity measure (i.e. Euclidean). The effect of k on the results is a key consideration. A larger k results in a smoother boundary effect and reduces the effect of noise. As the balance between classes becomes most disparaging, a larger k may not necessarily yield better results. If k is too large it will always predict the majority class. This particularly becomes an issue in unbalanced class problems, where the number of labels for one class is substantially larger than the other [HTF09] [Mur12] [HHKT16].

Selecting an optimal k is tied to the goals of our model and the cost of misclassification of each class should also be taken into account when selecting k. Measuring the effect of k on accuracy, $A = \frac{TP+TN}{TP+FP+FN+TN}$, presents an exponential relationship, with larger k values positively impacting overall performance. However, aggregating class-labelling performance of both classes in real-world applications is generally an incorrect approach.

The increase in accuracy is result of predicting the majority class. The cost of correctly classifying the majority class label is at a cost of misclassifying the minority class label, but the two are effectively cancelled out in aggregation [HV03].

The manner in which the distance between the unlabeled test data and training data is measured may be modified in implementation. Given two points in a feature space there are several possible distance measures including but not limited to the Euclidean, Minkowski , Chi-Square, and cosine similarity measure [MGU02] [MSD81] [HHKT16]. In a study comparing the performance of k-NN on the basis of these measurements, it was found that there was no clear single similarity measure that outperformed given categorical data, the Euclidean and Minkowski distance function performed best overall given numerical data, and the Chi-Square distance measure performed best for mixed-type data [HHKT16]. Based on these results, and the nature of the features to be derived from our supplied dataset, the ideal measure would be the Euclidean distance measure (Eq. 2.9). For further details of the other possible measures refer to [HHKT16].

$$d(x, y) = \sqrt{\frac{\sum_{i=1}^{m}(x_i - y_i)^2}{m}} \tag{2.9}$$

2.5.3 Decision Trees

Before we discuss conditional inference trees and random forests, we will first offer an introduction to decision trees. Decision trees may be employed for classification or regression on the basis of variable properties. The traditional structure presents the tree as inverted, with the root at the top and the leaves, or nodes at the bottom Figure 2.9. The tree begins with a single root node, which functions as the starting point. From here, the initial split into two or more branches occurs producing nodes. This process of splitting continues to occur until reaching a termination node, or leaf, containing the decision. Non-leaf nodes are associated with a decision question, which serves to produce the subsequent splits by determining the divisions. While a decision tree may split into

multiple branches, they can also be limited to splits of two, also referred to as binary decision trees. Binary splits is better in practice to limit complexity. We will be focusing on binary classification trees [Mwi13].

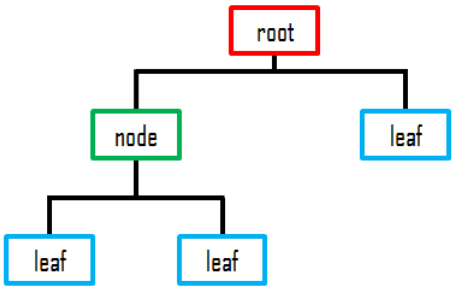


Figure 2.9: Decision tree structure

The root node contains the complete learning sample $\mathcal{L}$ at the top of the tree and every node can be described as a subset of the data. For a binary classification problem ($M = 2$), given an outcome variable $\mathbf{C} = \{C_1, C_2\}$ and a input feature vector $(X_1, \ldots, X_N)$, we seek to establish a relationship between the former and the latter. We can express this relationship using conditional probability, defined as $P(A|B)$, which is the probability of an event A occurring given that event B has already occurred. In the case of binary classification, the relationship between input features and class labels is $P(\mathbf{C} = C_M|X_1, \ldots, X_N)$. A node is a subset of $\mathcal{L}$, and may be terminal or non-terminal. Nonterminal nodes, or parent nodes, are those that will split into two daughter nodes. The splits are based on a Boolean condition on the value of an attribute variable, in response to a Yes/No data query i.e., Is the condition satisfied or not satisfied by the observed value? All observations contained in $\mathcal{L}$ meeting conditions for the specified variable drop down into the respective daughter node, with those that do not meet the condition dropping into the other. Terminal nodes are those that do not split, and are assigned a final class label. The stopping criteria may be caused by an absence of variables to split on (i.e. one observation in the node subset), or when further divisions do

not improve results. To use a tree for classification, a series of questions are asked in sequence, with the termination point, being the prediction. Given a data of unknown class label, attribute properties are used to travel through successive nodes of the tree, before finally assigning the class label corresponding to the terminal node. There may be more than one nonterminal node associated with the same attribute property, as well as more than one terminal node associated with the same class label. The complete set of all terminal nodes is referred to as the partition. The process of constructing a tree by the choice to split or not split each node into two daughter nodes is known as recursive partitioning. With recursive partitioning we seek to produce the terminal nodes that are homogeneous, although it should be noted that while complete homogeneity of terminal nodes is ideal, it is rarely achieved in real application.

For a supplied dataset there are countless possible splits. Not only must we make a decision on which attribute to split on the basis of, but also the manner of partition itself. With each split the future pool of possible splits decrease, however a future split may still occur using an attribute previously portioned on. For a nominal attribute there are $2^{r-1}-1$ possible splits where r is the number of levels [ZS10]. For an ordinal attribute of s distinct values, there are $s-1$ possible splits [ZS10]. Consequently, we require a means by which to measure goodness of a split. We seek to asses the homogeneities (impurities) of two daughter nodes [Ize08] [DHS12].

Our concern pertains to a binary outcome, thus for a node τ, the homogeneity of the node is defined by the probability $P\{\mathbf{C} = C_M|\tau\}$ for $M = 1, 2$. The most homogenous nodes are those with only one class of outcome such that the probability $P\{\mathbf{C} = C_M|\tau\} = 0 \quad \text{or} \quad 1$. The least homogenous nodes are those with equal likelihood of outcome of either class label such that the probability $P\{\mathbf{C} = C_M|\tau\} = \frac{1}{2}$. The impurity function is defined formally as $i(\tau) = \phi(P\{\mathbf{C} = C_M|\tau\})$ where ϕ is a function with the properties: (i) $\phi \geq 0$ and (ii) for $p \in (0, 1), \phi(p) = \phi(1-p)$ and $\phi(0) = \phi(1) < \phi(p)$. Some common choices of assessing node impurity include the Bayes

or the minimum error ($\phi(p) = \min(p, 1-p)$), the Gini index ($\phi(p) = p(1-p)$), and entropy ($\phi(p) = -p\log(p) - (1-p)\log(1-p)$) [where $0\log 0 = 0$] [ZS10].

An issue with decision tree is that with each split, future decisions for splits are made with a smaller sample. Terminal nodes may contain a subset too small to make inferences from. For this reason the tree may not be the best in terms of being representative of the data as a whole. Additionally, for most real-world applications, the constructed tree is too complex to be helpful. Pruning is a manner of improving accuracy and diminishing complexity, where a larger tree is generated initially before being reduced. To do so, it selects a subtree that offers the most predictive power and is least susceptible to noise. Decision trees efficiently present information, and handle outliers well, but there is a trade-off between the model's predictive performance and its simplicity [Ize08] [DHS12] [Mwi13].

2.5.4 Conditional Inference Tree

Although information measures are employed when deciding on the manner of splits in traditional decision trees, the actual statistical significance of the improvement is necessary to differentiate useful splits. For individual decision trees two key shortcomings are overfitting and variable selection bias. The former can be addressed by pruning, alternatively, [HHZ06] presented conditional inference trees to address the need for a statistical approach to recursive partitioning thats accounts for the distributional properties of the measures.

Conditional inference trees improve upon decision trees by correcting for issues in overfitting and feature selection biases through conditional distribution that measure the relationship amongst covariates and response variables [Ize08] [HHZ06] [Lev15]. They incorporate p-values into stopping and selection criteria and do not rely on post-hoc pruning [SKO14]. This method is biased towards variables that offer many possible splits. It aims to find the strongest association between the class label and attributes

and conducts tests for association using hypothesis tests. Selection of attributes and optimal split are independent of one another. Using binary recursive partitioning, the method begins by testing for association between each independent attribute and the response variable, before selecting the covariate with the strongest relationship. Similar to standard decision trees, a binary split is performed on this attribute, partitioning the dataset into two subsets using the binary predictor. The process is repeated for each subset until there are no longer any attributes associated with the outcome for a supplied statistical significance level using p-values. Attributes that will be partitioned are selected on the basis of a significance test as opposed to selecting the attribute that maximizes an information measure as is the method in standard decision trees [Ize08] [HHZ06] [Lev15].

2.5.5 Random Forests

A caveat of decision trees is that they can be an oversimplification model of the data. They can also be instable, as represented by minute perturbations to data (i.e. removal of a subset of the data) resulting in changes in the outcome. Individual trees may not be strong models, but can be improved when combined. Ensemble approaches seek to combat such issues and are less prone to instability. The random forest is a classifier made up of a collection of t tree structured classifiers generated from independent subsets of a given dataset. New data is classified using a majority vote of the ensemble of classifiers [ZW09][Bre01] .

Combining multiple decision trees to create random forests, counteracts this instability, and makes it robust to outliers. It also performs well with binary classification problems with unbalanced classes. Each of the individual trees are overfit, however random forests omit the pruning step that may be necessary in complex decision trees. Instead, it uses multiple independent trees that use different subsets and fit the data differently. For this reason, the random forest algorithm is comparatively less biased.

Random forests are especially appropriate when the input variables outweigh the observations, however it is also computationally intensive [Bre01].

The algorithm begins by selecting a randomly sampled subset of data (bootstrap sample). Recursive partitioning is performed using a randomly sampled selection of attributes at each split to generate complete trees. We can alternatively omit a random subset of attributes at each split, and use the complete set, a variation referred to as bagging [Bre01]. The process is performed for a specified number of trees to generate a forest with classification based on a majority vote of an ensemble of trees.

Selecting an optimal number of trees when constructing a forest seeks to produce a peak trade-off between the predictive performance and computation time. Due to the number of trees and their respective complexity, it is impractical to present them all. There is, in fact, no globally accepted method of tree selection, however there has been some work into seeking a viable option. [ZW09] explored mathematically selecting an optimal number of trees in random forest generation. For a constructed forest the accuracy following the removal of each tree is measured, in order to identify which has the minimal impact. For a forest F the accuracy is calculated as p_F. Each tree is represented by T. An $N-1$ approach is taken, and the accuracy of the forest ($p_{F_{-T}}$) is recomputed for (F_{-T}) following the removal of each tree T. The difference between the original accuracy and the accuracy following an excluded tree is computed for each T such that: $\Delta_{-T} = p_F - p_{F_{-T}}$. Following this $T^p = \arg\min_{T \in F}(\Delta_{-T})$ is computed to select the tree with the minimal difference. This paper explored the idea of starting with a random forest constructed with a large number of trees and gradually selecting trees to be excluded while maintaining performance. If a substantial increase in the number of trees is not reflected by significant improvement to classification, it is usually a better option to go with fewer trees.

[EAP11] advocated that reducing the variance of the tree ensemble selection method by cross-validation to improve performance. They specify improved performance is di-

rectly linked to diversity, which implies that in a good classifier the observations misclassified by each of the trees in the ensemble is unique to each tree. They presented a method to adaptively model the tradeoff between diversity and accuracy. Ensemble selection refers to the process of selecting individual classifiers from an original ensemble that produce accuracy on par with the original ensemble, but with reduced time and resource consumption. For an ensemble of size t trees, this involves examining a space of 2^{t-1}. Standard random forest implementation omit pruning, however in their paper they make a case for it. Selecting solely the most accurate individual classifiers to construct an ensemble, or considering only diversity for pruning are not the best. It would appear that the randomness reflected by the nomenclature of the algorithm can at times, work against it, with individual trees being deemed weak predictors due to random variable selection. The conclusion was that their method produced a relatively modest gain in performance, however they could draw no general conclusion on the optimal trade-off.

[OPB12] generated multiple random forests with number of trees increasing exponentially using base 2. They used Area Under Curve (AUC) of ROC curve, and percentage of features used in each to perform analysis of results. The takeaway of their analysis was a substantial number of trees may in some cases contribute significantly more to computational cost rather than performance. In particular, they found in their analysis that doubling the number of trees did not produce a significant difference (No significant difference between any result when number of trees is changed from 128 to 256, 512 ... 4096). However a significant change was seen at 64 trees relative to numbers preceding it. Thus they suggested 64-128 trees as a possible optimal range given their results. Evaluating the percentage of attributes used for classification, they saw an asymptotical tendency to utilize all attributes, with all attributes beings used with 128 trees. Therefore, an unwanted byproduct of random forests is a large forest may select unimportant attributes. Selecting an appropriate number of trees when conducting a random forest implementation necessitates consideration into a balance between significant performance

improvement and computation time.

2.5.6 Naive Bayes Classification

By the Bayes' Theorem we calculate posterior probability such that

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}.$$

This refers to the probability of event A occurring given that event B has occurred. Let $\mathbf{X}$ to be a n-dimensional vector containing feature data such that $\mathbf{X} = (X_1, ..., X_N)$. X_n refers to the n-th feature, and $\mathbf{C} = (C_1, ..., C_M)$ be a vector of class labels. C_m refers to the m-th class label. We then seek to use a Bayes' classifier to map the feature data to the class data. We can express Bayes' rule for our data as

$$P(C_m \mid \mathbf{X}) = \frac{P(\mathbf{X} \mid C_m)\, P(C_m)}{P(\mathbf{X})}.$$

For a test vector $\mathbf{X}$, we estimate the posterior probability $P(C_m \mid \mathbf{X})$ for each class m. To classify $\mathbf{X}$ we choose the class label with the largest estimated posterior probability such that

$$C = \arg\max_{C_m} \frac{P(\mathbf{X} \mid C_m)\, P(C_m)}{P(\mathbf{X})}.$$

However, estimating $P(\mathbf{X} \mid C_m)$ is computationally expensive in execution. Consider a discrete n-dimensional vector X with boolean variables. Estimating $P(C)$ is a simple task, however for $P(\mathbf{X} \mid C)$ we would need to estimate $2(2^n - 1)$ parameters. The Naïve Bayes Classifier is based on Bayes' theorem with a class conditional independence assumption. By this assumption, all predictor variables contribute independently to class labels and any dependence between input variables is disregarded. Consider an example given 2 features X_1 and X_2, thus

$$P(X_1, X_2 \mid C_m) = P(X_1 \mid X_2, C_m) P(X_2 \mid C_m).$$

which by conditional independence is equivalent to $P(X_1 \mid C) P(X_2 \mid C)$. The inclusion of the assumption prevents computational complexity associated with generating

the parameter vector. We can express the class conditional independence assumption mathematically for N features as

$$P(\mathbf{X} \mid C_m) = P(X_1, X_2, X_3, ...X_N \mid C_m) = \prod_{i=1}^{N} P(X_i \mid C_m).$$

For cases where the input features are continuous and class labels are discrete, a modified version is followed. In such cases it is assumed that $P(X_i \mid C_m)$ takes on a normal Gaussian distribution. Thus the Gaussian Naive Bayes (GNB) is used such that

$$P(X_i \mid C_m) = \frac{1}{\sqrt{(2\pi\sigma_{im}^2)}} \exp\Big[\frac{-(X_i - \mu_{im})^2}{2\sigma_{im}^2}\Big].$$

Refer to [Mit05] [JMN13] for more information.

2.5.7 Support Vector Machines

Support Vector Machines (SVM) is a supervised learning model that selects an optimal hyperplane to best separate two data classes. To do so it determines how to divide a space through use of decision boundaries during training. New unlabelled points can be tested for classification on the basis of where they fall relative to selected boundaries. For now we will focus on use of SVM on separable data. The identification of the best hyperplane is formed on the basis of the largest margin between two classes. The margin is the maximum width of the slab parallel to the hyperplane that contains no interior data points. The support vectors are the data points that are closest to the separating hyperplane. Thus the goal becomes to maximize the margin, implying an optimization problem.

Given data points x_i in n-dimensional space, we set corresponding classes y_i, where $y_i \in \{-1, 1\}$. Introduce the vector $\bar{w}$ to the minimization model, such that $\bar{w}$ is to be perpendicular to the median plane (or rather the hyperplane). We can then express each data points as a vector $\bar{x}$ from the origin. Then the projection of the vector $\bar{x}$ onto $\bar{w}$ is given by $\bar{w} \cdot \bar{x}$. The equation for any plane can be expressed as

$$\bar{w} \cdot \bar{x} - b. \tag{2.10}$$

Assume the two classes can be separated by a hyperplane given by $\bar{w}\cdot\bar{x}-b=0$. Equations of the two planes representing the margin given respectively by

$$\bar{w}\cdot\bar{x}-b=1 \quad \text{and} \quad \bar{w}\cdot\bar{x}-b=-1. \tag{2.11}$$

Introduce constraint to ensure data points remain on exterior of margin:

$$\bar{w}\cdot\bar{x}_i-b\geq 1 \quad \text{and} \quad \bar{w}\cdot\bar{x}_i-b\leq -1. \tag{2.12}$$

Given 2.12, we can combine these constraints by variable y_i where $y_i = \pm 1$ and is evaluated for each point x_i such that new constraint becomes

$$y_i\left(\bar{w}\cdot\bar{x}_i-b\right)\geq 1 \quad \text{with} \quad y_i=\pm 1. \tag{2.13}$$

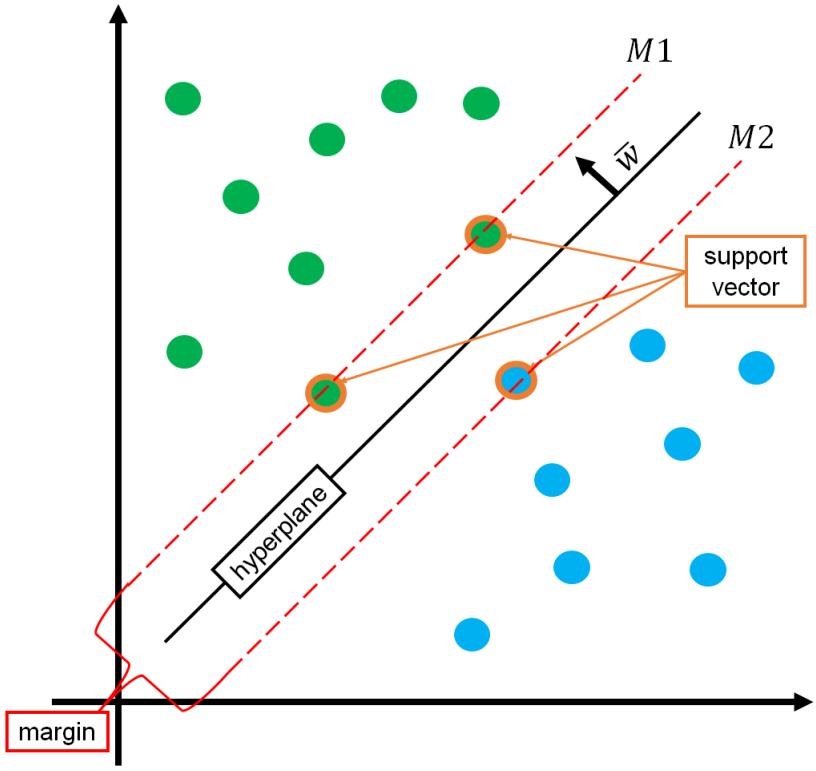


Figure 2.10: Support Vector Machines. Let the green points denote data with class 1 labels and the blue points denote data with class 0 labels. We seek to identify the optimal hyperplane separating the two data groups such that the margin is maximized. Support vectors are the points that exist closest to the hyperplane.

We want to maximize the margin (width between the two outlying planes). Consider margins $M1 = \bar{w} \cdot \bar{x}_+ - b = 1$ (associated with $y_i = 1$) and and $M2 = \bar{w} \cdot \bar{x}_- - b = -1$ (associated with $y_i = -1$) (refer to Figure 2.10). Then the margin is given by

$$\begin{aligned} margin &= \left(x_+ - x_-\right) \cdot \frac{\bar{w}}{\|\bar{w}\|} \\ &= \left(\left(\frac{1+b}{\bar{w}}\right) - \frac{-1-b}{\bar{w}}\right) \cdot \frac{\bar{w}}{\|\bar{w}\|} \\ &= \frac{2}{\|\bar{w}\|}. \end{aligned} \tag{2.14}$$

Thus we seek to maximize $\frac{2}{\|\bar{w}\|}$ which is equivalent to minimizing $\|\bar{w}\|$. Taking the problem to be $\min \frac{1}{2}\|\bar{w}\|^2$ for simplicity with later steps. Therefore the optimization problem to maximize the margin is given by

$$\min \frac{1}{2}\|\bar{w}\|^2 \quad \text{subject to} \quad y_i\left(\bar{w} \cdot \bar{x}_i - b\right) \geq 1. \tag{2.15}$$

We can use Lagrange Multipliers to rewrite the problem without using constraints as

$$L = \frac{1}{2}\|\bar{w}\|^2 - \sum_i \alpha_i \left[y_i\left(\bar{w} \cdot \bar{x}_i - b\right) - 1\right]. \tag{2.16}$$

Taking the partial derivative with respect to unknown components ($\bar{w}$ and b) and equating to 0 respectively produces

$$\frac{\partial L}{\partial \bar{w}} = \bar{w} - \sum_i \alpha_i y_i \bar{x}_i = 0 \tag{2.17}$$

and

$$\frac{\partial L}{\partial b} = -\sum_i \alpha_i y_i = 0. \tag{2.18}$$

Equation 2.17 implies

$$\bar{w} = \sum_i \alpha_i y_i \bar{x}_i \tag{2.19}$$

and Equation 2.18 implies

$$\sum_i \alpha_i y_i = 0. \tag{2.20}$$

Substituting these expressions (Eq. 2.19 and Eq. 2.20) into Eq. 2.16 we can rewrite L such that

$$\begin{aligned} L &= \frac{1}{2}\|\bar{w}\|^2 - \sum_i \alpha_i \left[y_i \left(\bar{w} \cdot \bar{x}_i - b \right) - 1 \right] \\ &= \frac{1}{2}(\bar{w}_i) \cdot (\bar{w}_j) - \sum_i \alpha_i y_i \bar{x}_i \cdot (\bar{w}_j) - \sum_i \alpha_i y_i b + \sum_i \alpha_i \\ &= \frac{1}{2}\left(\sum_i \alpha_i y_i \bar{x}_i \right) \cdot \left(\sum_j \alpha_j y_j \bar{x}_j \right) - \left(\sum_i \alpha_i y_i \bar{x}_i) \cdot \left(\sum_j \alpha_j y_j \bar{x}_j \right) - \sum_i \alpha_i y_i b + \sum_i \alpha_i \right. \\ &= \frac{1}{2}\left(\sum_i \alpha_i y_i \bar{x}_i \right) \cdot \left(\sum_j \alpha_j y_j \bar{x}_j \right) - \left(\sum_i \alpha_i y_i \bar{x}_i) \cdot \left(\sum_j \alpha_j y_j \bar{x}_j \right) + \sum_i \alpha_i \right. \\ &= \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i y_i \alpha_j y_j \bar{x}_i \cdot \bar{x}_j \end{aligned} \tag{2.21}$$

Therefore the problem can be seen as maximizing $L = \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i y_i \alpha_j y_j \bar{x}_i \cdot \bar{x}_j$ subject to $\sum_i \alpha_i y_i = 0$ and $\alpha_i \geq 0$ for $i = 1, ...n$. This produces the equation for our training. Note that the optimization depends only on dot product of pairs of samples. Recalling the decision rule we initially had (Eq. 2.13), and substituting in our expression from 2.19 results in

$$y_i \left(\sum_i \alpha_i y_i \bar{x}_i \cdot \bar{x} - b \right) \geq 1 \quad \text{with} \quad y_i = \pm 1. \tag{2.22}$$

The implication is that the decision rule depends on pairs of samples. Thus far we have examined the problem necessary to run the training portion of algorithm in order to deduce the optimal hyperplane. Classification is a simple process of considering new points and examining where they fall relative to the hyperplane. SVM is a primarily two class classifier, however multi-class problems can be performed by a one versus all approach. One would separate one class for each N classes and combine to form a multi-class solution. We were considering linearly separable cases for developing the model. For nonlinearly separable cases we would need to perform a transformation to take the data to a higher dimension where a linear hyperplane could be selected for separation. Take this transformation to be ϕ. Since our maximization only depends on dot products

we can transform our dot product $\phi(\bar{x}_i) \cdot \phi(\bar{x})$ which we equate to some kernel function $K(\bar{x}_i, \bar{x})$. This transformation takes the data into a more convenient space. It is not necessary to know the transformation itself. Selecting an appropriate kernel function can be a difficult task. There are many possible kernel functions such as polynomial kernels, and the Gaussian Radial Basis function kernel [HDO$^+$98] [SC08] [SS02].

2.6 Introduction to Textural Analysis

Texture is a significant visual input and while easy to recognize distinctions in the feature visually, it can be challenging to define. Physical surface qualities can manifest themselves as texture. Comparing healthy and non-healthy brain tissue on MRI, there are variations in intensity in both. Thus, finding the recurring patterns within these variations becomes the objective as a means of differentiating the two. In order to recognize texture in images we seek to examine spatial variation in gray level values in order to explore texture properties [TJ+93]. To quantify texture we can employ Gray-Level Co-occurrence Matrices (GLCM) to produce textures features as first proposed by [HS+73]. A co-occurrence matrix, or co-occurrence distribution, expresses the frequencies of co-occurring values over a specified offset of adjacent locations. Using a gray-scale image a GLCM can be derived. A co-occurrence matrix is generated for each voxel using some specified surrounding region about each of the voxels and accounting for intensity pairings (Figure 2.11). From these co-occurrence matrices, features can be computed that are mapped back to voxel positions [ACB+03]. The co-occurrence matrix can divulge details about spatial distribution of gray levels. In medical imaging, automated extraction of texture features for use in classification processes to identify abnormal tissue is a well-explored application [TJ+93].

Given an 2-D input image $I(x, y)$, consider a patch of size $M \times N$ taken from image I. Note that the intensity values of this patch can be represented by f such that $f(m, n)$ is the intensity at sample row m, column n of specified patch (corresponds to sampling image $I(x, y)$ within the patch). Let the displacement distance vector (the offset) be $d = (\triangle x, \triangle y)$. Let the number of gray levels, the scale factor for intensity values of the image I, be represented by G. The co-occurrence matrix generated will be of size $G \times G$. Binning of intensity values to reduce gray levels is a standard protocol, to avoid issues to temporary data storage with images with a large number of unique intensity values. We take the co-occurrence matrix to be $P[i, j]$ (i, $j \in \{1, ..., G\}$). For each entry of $P[i, j]$,

counts number of times that pairs of gray levels i and j occurs in pixels separated by distance d. Thus

$$P[i,j] = \left[\frac{1}{(M-\triangle x)(N-\triangle y)}\right] \sum_{n=1}^{N-\triangle y} \sum_{m=1}^{M-\triangle x} A(m,n)$$

$$A(m,n) = \begin{cases} 1, & \text{if } f(m,n) = i \quad \text{and} \quad f(m+\triangle x, n+\triangle y) = j \\ 0, & \text{otherwise.} \end{cases}$$

Note that the elements of $P[i,j]$ are normalized, thus in reality GLCMs contain probabilities as opposed to counts [ZRP+11] [TJ+93] [JKS95] [A+08]. The offset can alternatively be represented using the distance and the angle. However, a single GLCM may not be always serve as a sufficient descriptor. Solely using a horizontal offset may lack sensitivity to textural attributes present in the vertical direction. Thus it is standard practice to produce multiple GLCMs using an array of offsets and co-occurrence matrices can be recomputed for multiple displacements. A possible way of doing so is by generating symmetric GLCMs. As previously defined, the co-occurrence matrix is not symmetric, however symmetric matrices as described in Haralick's original paper ([HS+73]) may be computed (Refer to Figure 2.11). Computing symmetric matrices allows discretized directions to be reduced. In 2-D, the 8 discretized directions can be collapsed to 4 possible spatial directions [EDA+15]. Consider a specified pixel distance of D, a possible method of generating a symmetric co-occurrence matrix from a 3-Dimensional neighbourhood of a given volume is to sample from D units in each direction. Thus there are 26 discretized directions which can be collapsed to 13 spatial directions [ACB+03]. Thus, we can express the offsets using this method in 3-Dimensions as

$$[0,D,0],\ [-D,D,0],\ [-D,0,0],[-D,-D,0],\ [-D,D,-D],\ [-D,0,-D],$$
$$[-D,-D,-D],\ [0,-D,-D],\ [D,-D,-D],\ [D,0,-D],\ [D,D,-D],\ [0,D,-D],$$
$$[0,0,-D].$$

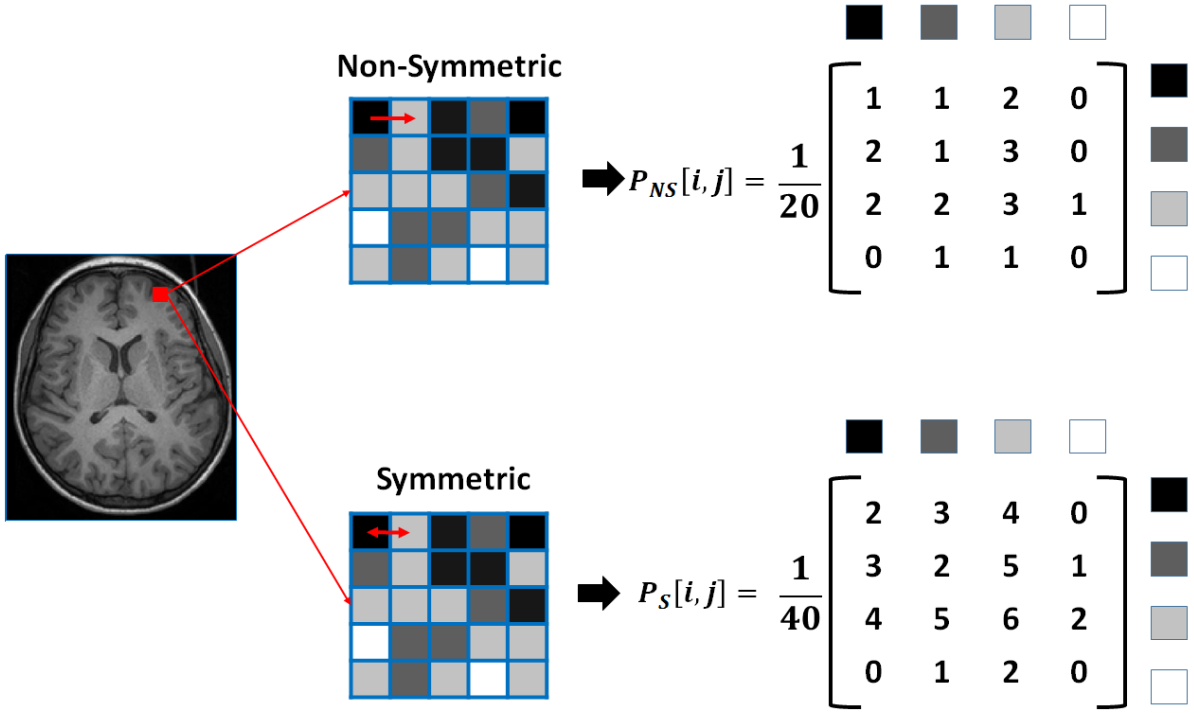


Figure 2.11: Gray Level Co-occurrence Matrix (GLCM) generated in 2-D. Given a region of interest, a GLCM can be produced by tallying the instances of voxel intensity pairings separated by a specified offset (Distance between pixel of interest and neighbour). Consider an area selected from an axial slice of a T1-weighted image. We seek to generate a Non-symmetric GLCM, and a Symmetric GLCM for the same region, using $G = 4$ gray levels, indicating that our resulting GLCMs will be 4 by 4 matrices. We will employ an offset of [0,1], which implies a shift of 1 to the right. To generate a symmetric matrix we count a pairings of the same levels in both directions. We then examine the occurrence of each pairing of levels using the defined offset and fill in the array. Finally the GLCM requires normalization, done by dividing each component by the sum off values.

From the co-occurrence matrices we can compute some possible textural features as follows [HS+73] [ACB+03], [A+08]:

Angular Second Momentum	$\sum_{i=1}^{G}\sum_{j=1}^{G}\{P[i,j]\}^2$
Entropy	$-\sum_{i=1}^{G}\sum_{j=1}^{G}P[i,j]log\{P[i,j]\}$
Correlation	$\sum_{i=1}^{G}\sum_{j=1}^{G}\frac{ijP[i,j]-\mu_x\mu_y}{\sigma_x\sigma_y}$
Contrast	$\sum_{n=0}^{G-1}n^2\{\sum_{i=1}^{G}\sum_{j=1}^{G}P[i,j]\} \quad , \quad \lvert i-j\rvert = n$
Homogeneity	$\sum_{i=1}^{G}\sum_{j=1}^{G}\frac{P[i,j]}{1+(i-j)^2}$
Variance	$\sum_{i=1}^{G}\sum_{j=1}^{G}(i-\mu)^2P[i,j]$
Sum Mean	$\frac{1}{2}\sum_{i=1}^{G}\sum_{j=1}^{G}(iP[i,j]+jP[i,j])$
Inertia	$\sum_{i=1}^{G}\sum_{j=1}^{G}\{i-j\}^2P[i,j]$
Cluster Shade	$\sum_{i=1}^{G}\sum_{j=1}^{G}\{i+j-\mu_x-\mu_y\}^3P[i,j]$
Cluster Tendency	$\sum_{i=1}^{G}\sum_{j=1}^{G}\{i+j-\mu_x-\mu_y\}^4P[i,j]$
Max Probability	$\max_{i,j=1}^{G}\{P[i,j]\}$
Inverse Variance	$\sum_{i=1}^{G}\sum_{j=1}^{G}\frac{P[i,j]}{(i-j)^2} \quad , \quad i \neq j$

Table 2.5: Textural features

with $P_x(i) = \sum_{j=1}^{G}P[i,j]$, $P_y(j) = \sum_{i=1}^{G}P[i,j]$, μ_x = mean of P_x, σ_x = standard deviation of P_x, μ_y = mean of P_y, σ_y = standard deviation of P_y, and μ = mean value of P.

The attributes are reflective of different traits in the image, which in the context of lesion identification, can contribute meaningful information to analysis. Angular Second Momentum (ASM) is a strong measure of uniformity. A homogeneous region will contain a small range of gray levels, producing a GLCM composed of few non-zero values, albeit high ones. Thus, the sum of squares of such a case would be high [A+08]. Entropy acts as a measure of randomness in intensity. Low measures occur with inhomogeneous regions, while high entropy implies homogeneity [A+08] [GW02]. The Correlation exam-

ines gray level linear dependence of considered voxels [A$^+$08]. Contrast considers local intensity variation by examining differences between neighboring voxels. The contrast value increases with higher magnitudes of variation. It favours contributions of $P[i, j]$ for which $i \neq j$, thus a contrast of 0 indicates no difference between a reference voxel and a neighbour voxel [A$^+$08] [HS$^+$73]. Homogeneity, also known as the inverse difference moment, is the opposite of contrast, with high values indicting the presence of smooth texture with minimal variation. On the other hand, low values imply an inhomogeneous images while relatively higher values occur with homogeneous images [A$^+$08] [HS$^+$73]. Elements differing from the average value are weighted higher with Variance. The opposite is true for Inverse Variance [A$^+$08]. The Sum Mean is simply the mean sum of internship of adjacent voxels to a central voxel [HS$^+$73]. Inertia values are weighted based on the difference between i and j, with the weight increasing exponentially as $(i - j)$ increases [HB07]. Cluster shade and cluster tendency evaluate the symmetry of the matrix (skewness). Higher values imply a lack of symmetry [HS$^+$73] [PVL$^+$14]. Finally, Max Probability is the maximum element of $P[i, j]$, as a means of evaluating for the dominant pixel pairings occurring [HB07]. These features can be computed using *cooc3d* in MATLAB.

Chapter 3

Experiment and Results

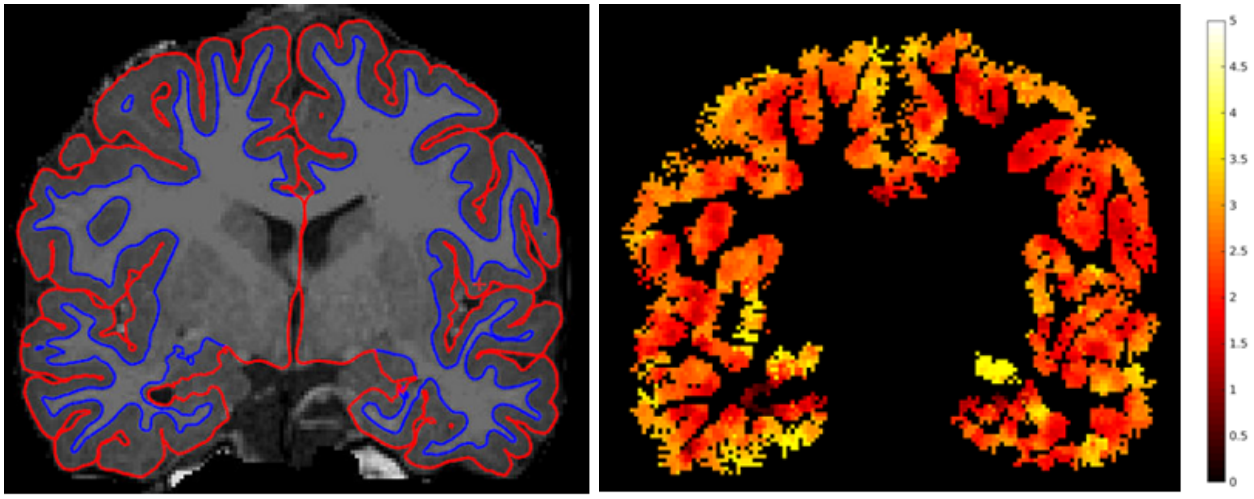
3.1 Preliminary Morphometry Experiments

3.1.1 Identifying Epileptogenic Lesions using Cortical Thickness Mappings

Abnormal cortical thickness is one of the features of FCD and one of the most fundamental for brain imaging [LWS13]. The aim of this experiment was to determine if using cortical thickness as an objective measure could assist with identifying subtle FCD on volumetric T1-weighted MRI. Many previous publications in the field have used volumetric T1 images due to their thinner slices. Of the 31 MRI-Positive patients, 5 were known to have focal cortical thickening that was visible on visual inspection of the MRI by a neuroradiologist prior to surgical intervention. These were cases NL008, NL0021, NL0023, NL0036, and NL039. Control cases were omitted for this experiment.

Cortical thickness measurements were acquired using the FreeSurfer software. FreeSurfer constructs models to represent the pial surface (the boundary between gray matter and cerebrospinal fluid) and gray-white matter junction surface with the surfaces generated being in the form of triangular meshes (Figure 3.1a, 3.2, and 3.3). Once the pial surface and gray-white matter junction surfaces are established, additional anatomical mea-

sures may be computed including cortical thickness and sulcal depth. The surfaces have point-to-point correspondence, and the distance between these two surfaces expresses the cortical thickness. Thus to compute the cortical thickness, first the shortest distance from the gray-white matter junction to the pial surface is computed. As the surfaces are deformable, this is done by finding the closest point on the opposite surface and expanding each of the vertices of the inner surface to the outer surface. The same is then done in reverse, taking the measurements of the shortest distance from the pial surface to gray-white matter surface. The average of these two values is taken to be the cortical thickness for a given vertex (Figure 3.1b). Cortical thickness measures are saved as a vector of values by FreeSurfer. For each case, these thickness vectors were sampled into a volume that could be read in MATLAB.

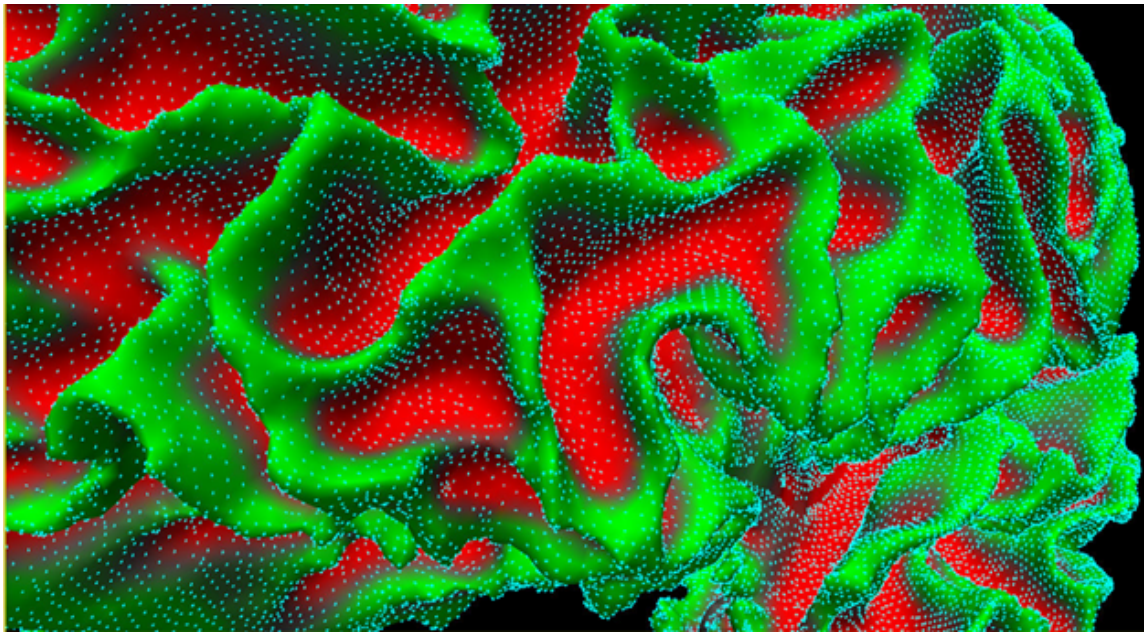


(a) (b)

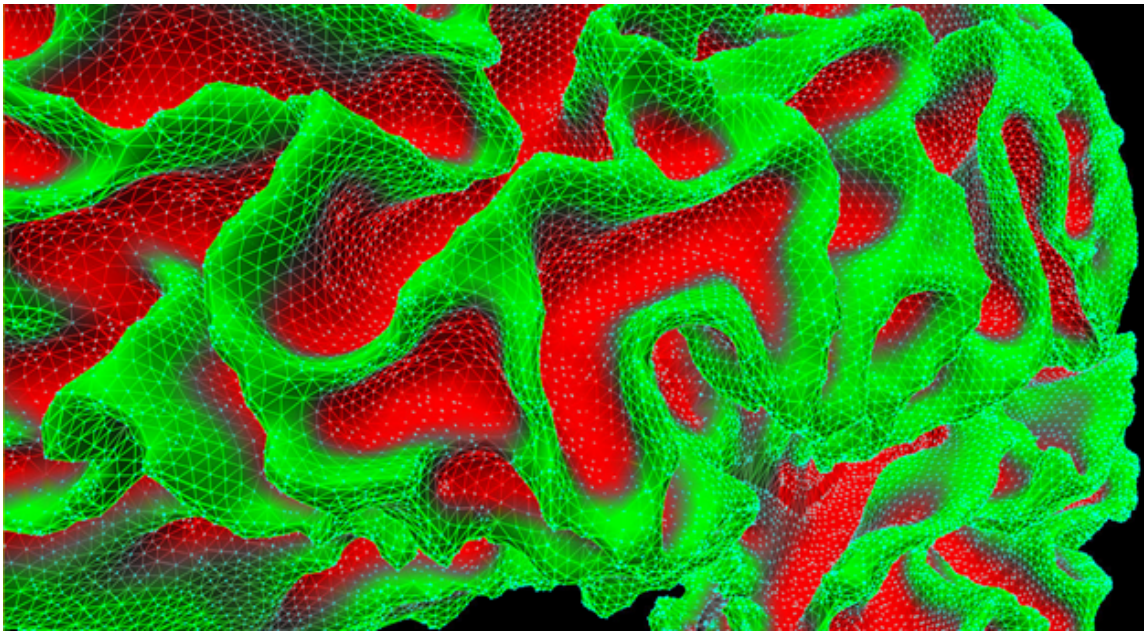
Figure 3.1: Cortical thickness generation. (a) Depicts a T1 axial slice of case NL021 with the pial surface and gray-white matter junction surface respectively indicated in red and blue. (b) Shows the axial slice of the corresponding cortical thickness volume produced by backprojecting the surface map to the individual volumes space.

FreeSurfer surfaces may also be inflated to improve visualizations of features as seen in Figure 3.4. Visualizing the cortical volumes indicates that some processing is necessary to improve detection. To do so, we used Gaussian filtering followed by gamma filtering. Gaussian Filtering is used in image processing to blur images using the Gaussian function, usually as a means of reducing artifacts or noise and smoothing images. In 1-dimension, the zero-mean Gaussian kernel is given by $G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$ where σ refers to the standard deviation. In 2-dimensions, as is the case with a single image, the function is the product of the Gaussian kernel in both directions, and can be represented by $G(x) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$. Similarly in 3-D the Gaussian kernel is given by $G(x) = \frac{1}{(\sqrt{2\pi\sigma^2})^3} e^{-\frac{x^2+y^2+z^2}{2\sigma^2}}$. The term $\frac{1}{(\sqrt{2\pi\sigma^2})^3}$ acts as a normalization factor and guarantees the average gray level of the image remains the same following application of the filter. The filter is non-uniform with voxels located at the center being weighted higher than those at the margin. The Gaussian kernel is isotropic regardless of dimension. In 1-D it can be roughly described as a bell, in 2-D as a circle, and in 3-D as a sphere [Rom08]. For an MRI volume 3-D Gaussian filtering can be performed in MATLAB using the imgaussfilt3 function which takes on additional arguments for Filter size and σ, with Filter size referring to the size of the discretized Gaussian filter. In 1-D altering the σ, impacts the width of the Gaussian and can be described as a scaling parameter. Thus in 2-D and 3-D changes to σ affect the blur, and larger values of sigma produce a greater blur, as illustrated in Figure 3.5.

A gamma filter, or gamma correction can induce a nonlinear contrast adjustment to an image. The gamma filter applied to image $I(x)$ is $J(x) = T[(I(x)] = S \cdot I(x)^{\gamma}$ for constant values S and γ. When $S = 1$ and $\gamma = 1$, an identity transformation occurs. For $\gamma < 1$ darker intensities are enhanced, and for $\gamma > 1$, brighter intensities are enhanced (refer to Figure 3.6). Figure 3.7 demonstrates the impact of γ.

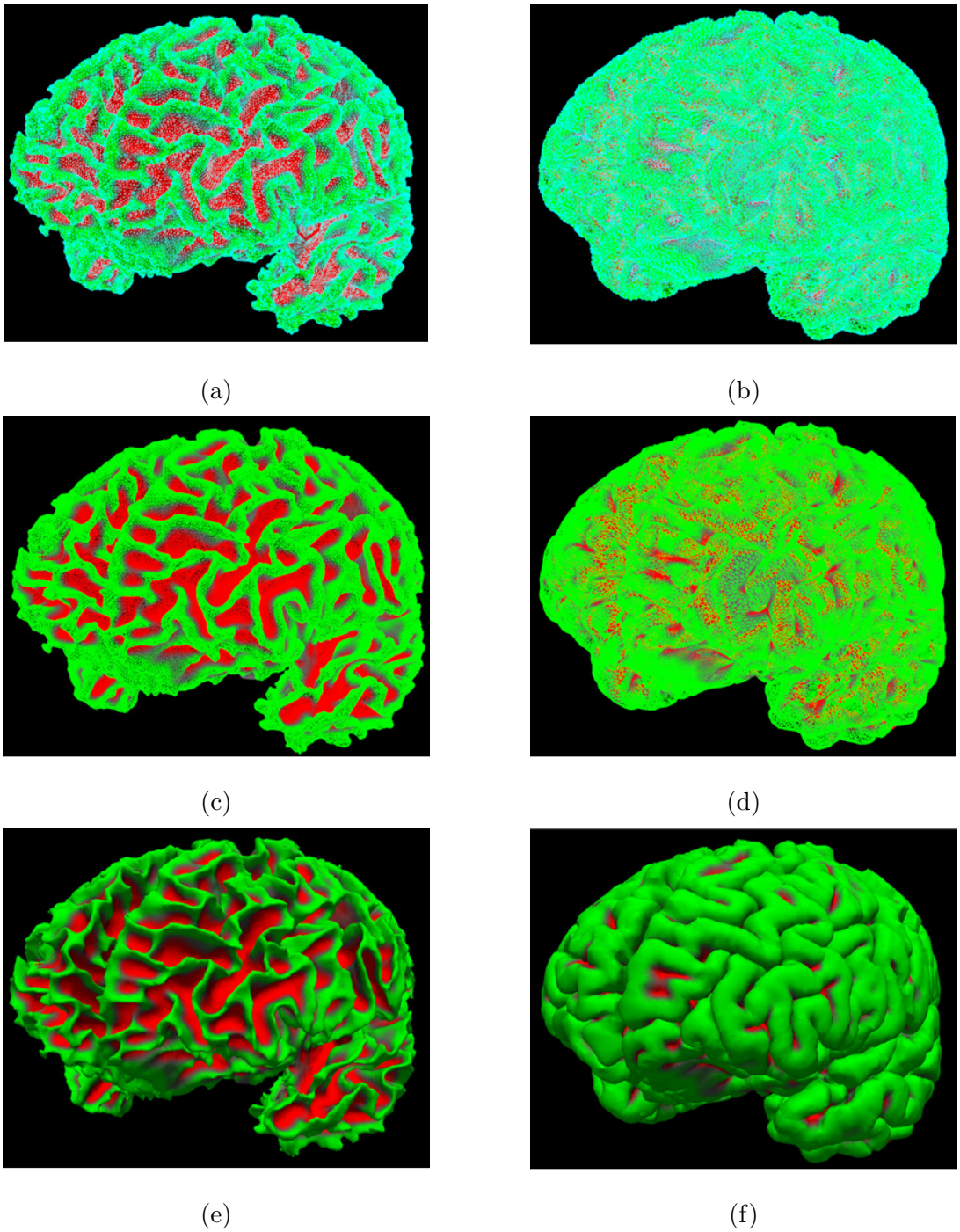


(a)



(b)

Figure 3.2: FreeSurfer surface representation. The gray-white matter junction of Case NL021 represented as a surface and in the form of triangular meshes. (a) Close up of gray-white matter junction surface with vertices (b) Close up of gray-white matter junction represented as a mesh.

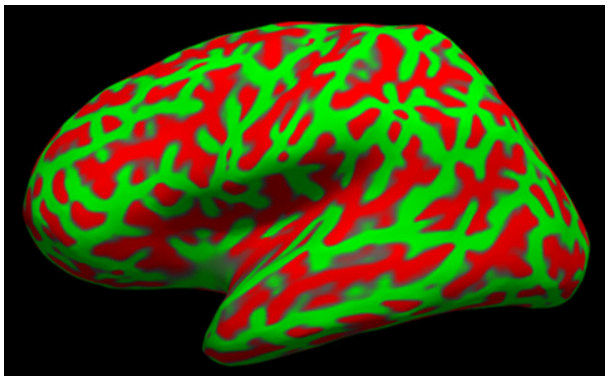


(a) (b)

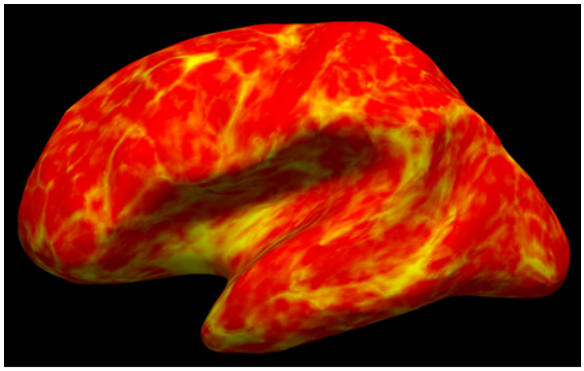
(c) (d)

(e) (f)

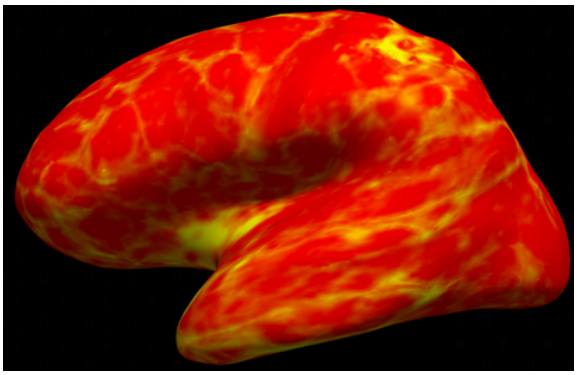
Figure 3.3: Pial and gray-white matter junction surfaces. (a) Gray-white matter junction mesh with vertices, (b) Pial matter junction mesh with vertices, (c) Gray-white matter junction mesh, (d) Pial matter junction mesh, (e) Gray-white matter junction surface, and (f) Pial matter junction surface



(a)

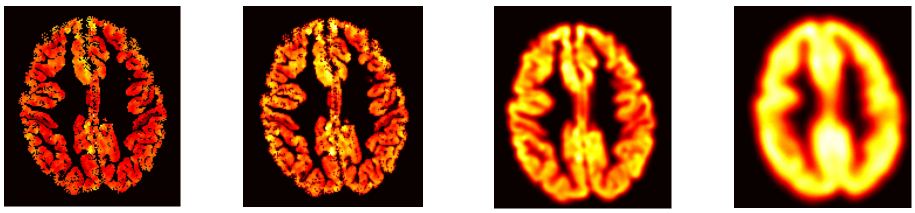


(b)



(c)

Figure 3.4: FreeSurfer inflated volumes: (a) Inflated MRI-Positive subject NL001 (red regions are sulci and green regions are gyri), (b) Cortical thickness values backprojected onto inflated volume of control NL030, and (c) Cortical thickness values backprojected onto inflated volume of MRI-Positive Case NL039. Surfaces can be inflated or flattened to aid in visualizing features, in this case cortical thickness. Comparing the control NL030 to the MRI-Positive subject NL039, we can note that both in healthy cases and FCD cases natural variations are present. There does appear to be a thicker region as indicated by the bright spot at the top of NL039. This does in fact correspond to the approximate location of the true lesion, however, further processing must be performed to improve detection.



(a) (b) (c) (d)

Figure 3.5: Effect of σ on Gaussian filter. Applied 3-D Gaussian filter using filter size of 17 and various σ values:(a) Original, (b) $\sigma = 0.5$, (c) $\sigma = 1.9$, and (d) $\sigma = 4.5$. Displayed axial slices.

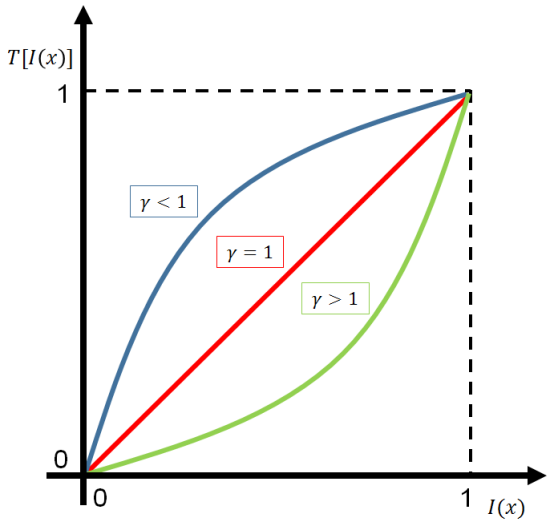
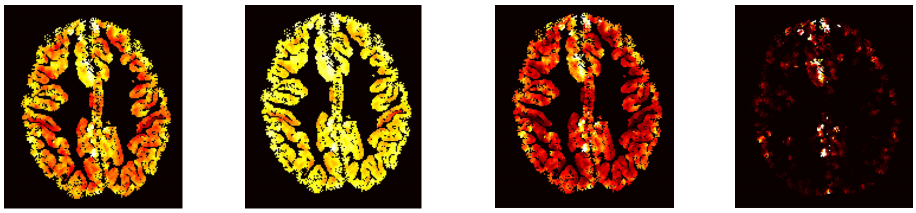


Figure 3.6: Gamma filtering on intensity of image I(x) for different values of γ. Intensity of image is normalized between 0 and 1.



(a) (b) (c) (d)

Figure 3.7: Effect of γ on gamma filter. Applied 3-D gamma filter for various γ values: (a) Original, (b) $\gamma = 0.5$, (c) $\gamma = 2$, and (d) $\gamma = 9$. Displayed axial slices.

Thus, for each case a 3D-Gaussian smoothing filter (sigma=1.9, filter size=17) was applied to the cortical volume followed by a gamma correction. A threshold was used on intensity values of vertices to omit points that fell outside the top 5%. The selected vertices were matched to their structure name and hemisphere using annotation files produced by FreeSurfer during pre-processing. FreeSurfer performs this structure labelling using a probabilistic atlas mapped to a common space. Once selected vertices were labelled, a secondary threshold of 5 mm was applied to the cortical thickness values. This was the standard upper limit in Freesurfer, however these values were initially included to allow for vertex matching.

For preliminary experiments we will consider coarse labelling and fine labelling. Coarse labelling corresponds to the brain being labelled to 8 possible locations, 4 per hemisphere: right and left frontal, temporal, parietal, and occipital. Fine labelling corresponds to the brain being labelled to 70 possible locations, 35 for each hemisphere (Refer to Table 2.1, Figure 2.3).

The resulting clusters of abnormal cortical thickness were first classified using coarse labels. The location of clusters was compared to surgical resection site on post-operative Computed tomography (CT) or T1-weighted MRI scans. For the 31 patients, using coarse labelling to the 8 major lobes resulted in clusters of abnormal cortical thickness identified in all 31 patients and in 22 patients these locations corresponded to surgical resections sites. Subjectwise sensitivity acts as our key performance evaluator, with lesional sensitivity and specificity serving to offer supplementary information. For preliminary morphometric experiments, the subjectwise specificity cannot be computed due to control cases not being included. The overall lesional sensitivity of cortical thickness measures to identify FCD was 69% and the lesional specificity was 55%. The overall subjectwise sensitivity was 71%. Of the 5 patients that had focal cortical thickening by visuals inspection, objective measures of cortical thickness showed that the lesional sensitivity and lesional specificity were 100% and 74% respectively and subjectwise sensitivity

was also 100%. These results are summarized in Table 3.1.

	Overall(%)	Cases Identified to show Focal Cortical Thickening at Visual Inspection (%)
Lesional Sensitivity	71	100
Lesional Specificity	55	74
Subjectwise Sensitivity	69	100

Table 3.1: Results of coarse labelling analysis - sensitivity and specificity considering all cases versus solely considering cases identified during visual inspection to present focal cortical thickening

The next step was to examine whether the lesion location could be further isolated. Thus, the overall process was repeated using the finer labelling. Vertex location was selected in the same manner as performed using coarse labelling, with a modification at the lobe matching stage. Instead of matching selected vertices to the 8 major lobes, we now match to the 70 sub-regions. The complete annotation volume produced by FreeSurfer during pre-processing (the left hemisphere is presented in Figure 2.3) was used to identify the sub-regions such that a cluster could fall into 70 possible locations. Among the 31 patients, using fine labelling resulted in clusters of abnormal cortical thickness identified in all patients and in 16 patients these locations corresponded to surgical resection sites. The overall lesional sensitivity of cortical thickness measures to identify FCD was 33% and the lesional specificity was 85%. The overall subjectwise sensitivity was 52%. Considering the 16 patients which were correctly colocalized, the lesional sensitivity was 59% and lesional specificity was 86%. From the 5 patients that had focal cortical thickening by visuals inspection, cortical thickness measurements showed that the lesional sensitivity and lesional specificity were 78% and 89% respectively. Subjectwise sensitivity was 80% with only case NL023 not being colocalized. These results are summarized in Table 3.2, and selected results are shown in Figure 3.8.

	Overall (%)	Cases identified to show focal cortical thickening at visual inspection (%)
Lesional Sensitivity	33	78
Lesional Specificity	85	89
Subjectwise Sensitivity	52	80

Table 3.2: Results of fine labelling analysis - sensitivity and specificity considering all cases versus solely considering cases identified during visual inspection to present focal cortical thickening

An issue with the method was low specificity, implying a high number of false positives. However this is not evident when comparing corresponding slices to resected regions (Figure 3.8). The reason for the high false positives is that the method ends up identifying clusters in areas that possess higher cortical thickness in all subjects. Consider Figure 3.9 which depicts the cluster selections of subject NL039. The implication is that the method would work best as a supplementary tool for verification.The complete pipeline of the employed method is outlined in Figure 3.10.

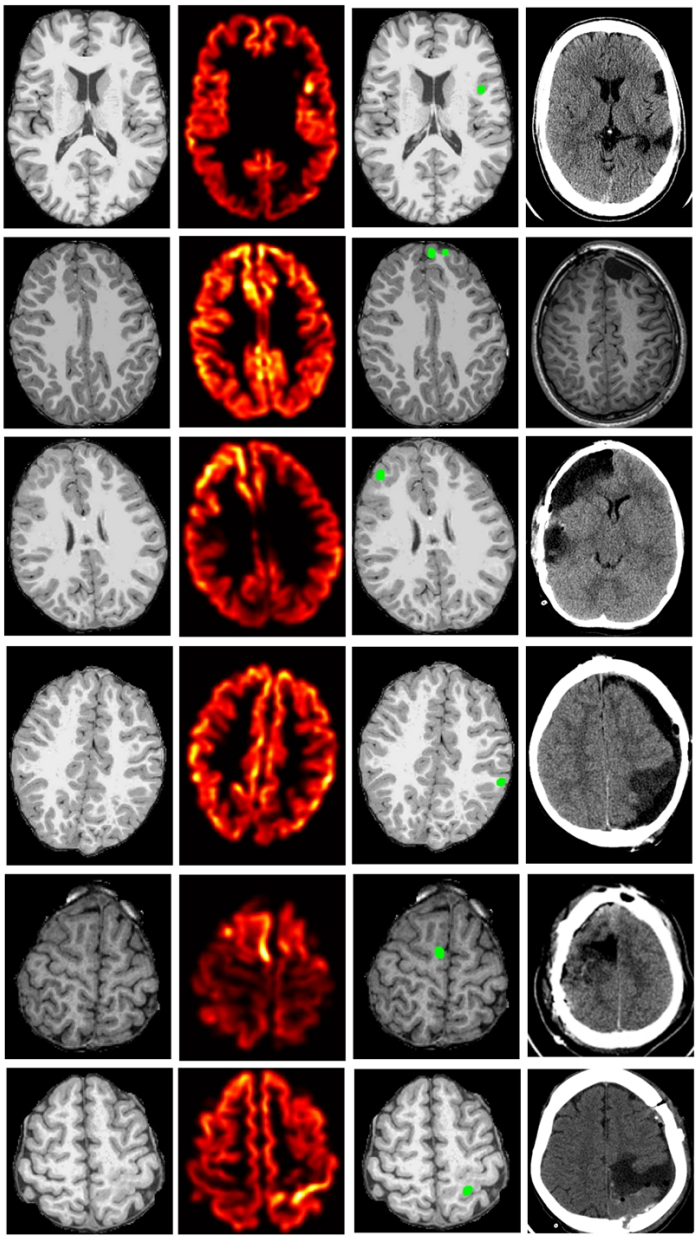
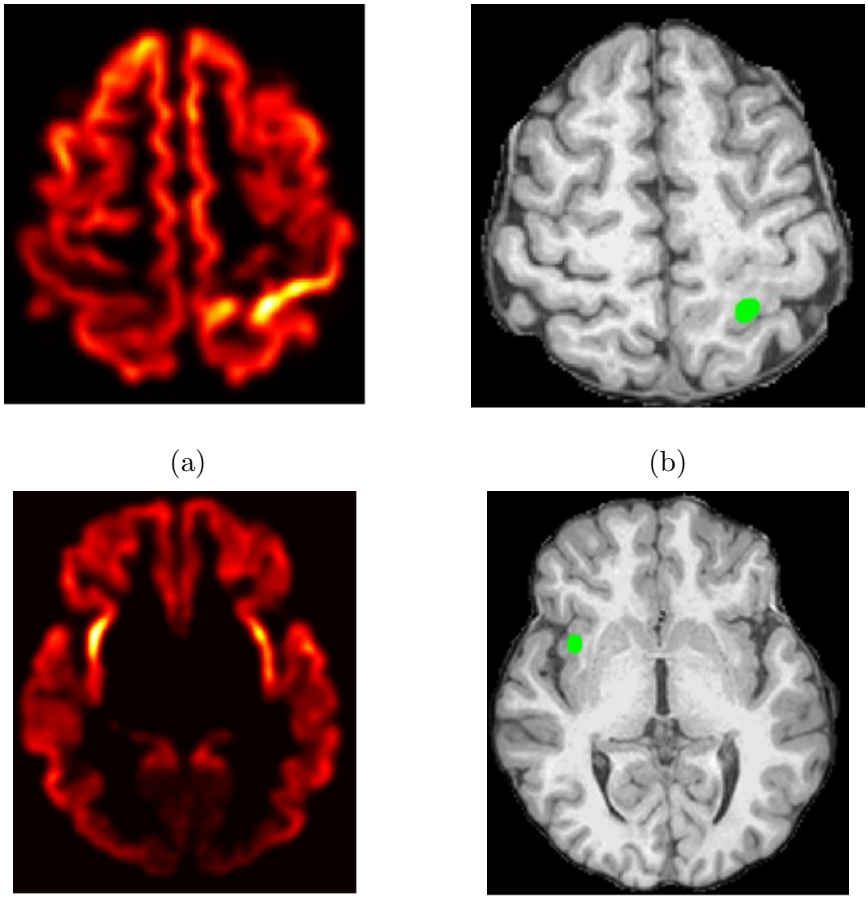


Figure 3.8: Selected results of volumetric T1 MRI processing using cortical thickness. (Column 1) Original MRI image after normalization and skull stripping , (Column 2) Showing cortical thickness measurements, followed by the application of Gaussian Smoothing and Gamma filtering. Visible cluster of abnormal signal indicates region of lesion. (Column 3) Original image with green overlay indicating vertices identified by program to indicate lesion, (Column 4) Surgical resection site on post-operative CT or T1 scan. Scan indicating location of lesion was removed



(a) (b)

(c) (d)

Figure 3.9: Comparison of TP and FP selections in subject NL039. (a) Subject NL039 cortical thickness measurements, followed by the application of Gaussian Smoothing and Gamma filtering. Visible cluster of abnormal signal indicates region of lesion. (b) Green overlay indicating vertices identified by program to indicate lesion. Corresponds to location of true lesion. (c) Cortical thickness measurements, followed by the application of Gaussian Smoothing and gamma filtering of alternate axial slice of same subject. Visible cluster of abnormal signal indicates program selection for region of lesion. (d) Alternate axial slice with green overlay indicating vertices identified by program to indicate lesion. Location does not correspond to true resected area.

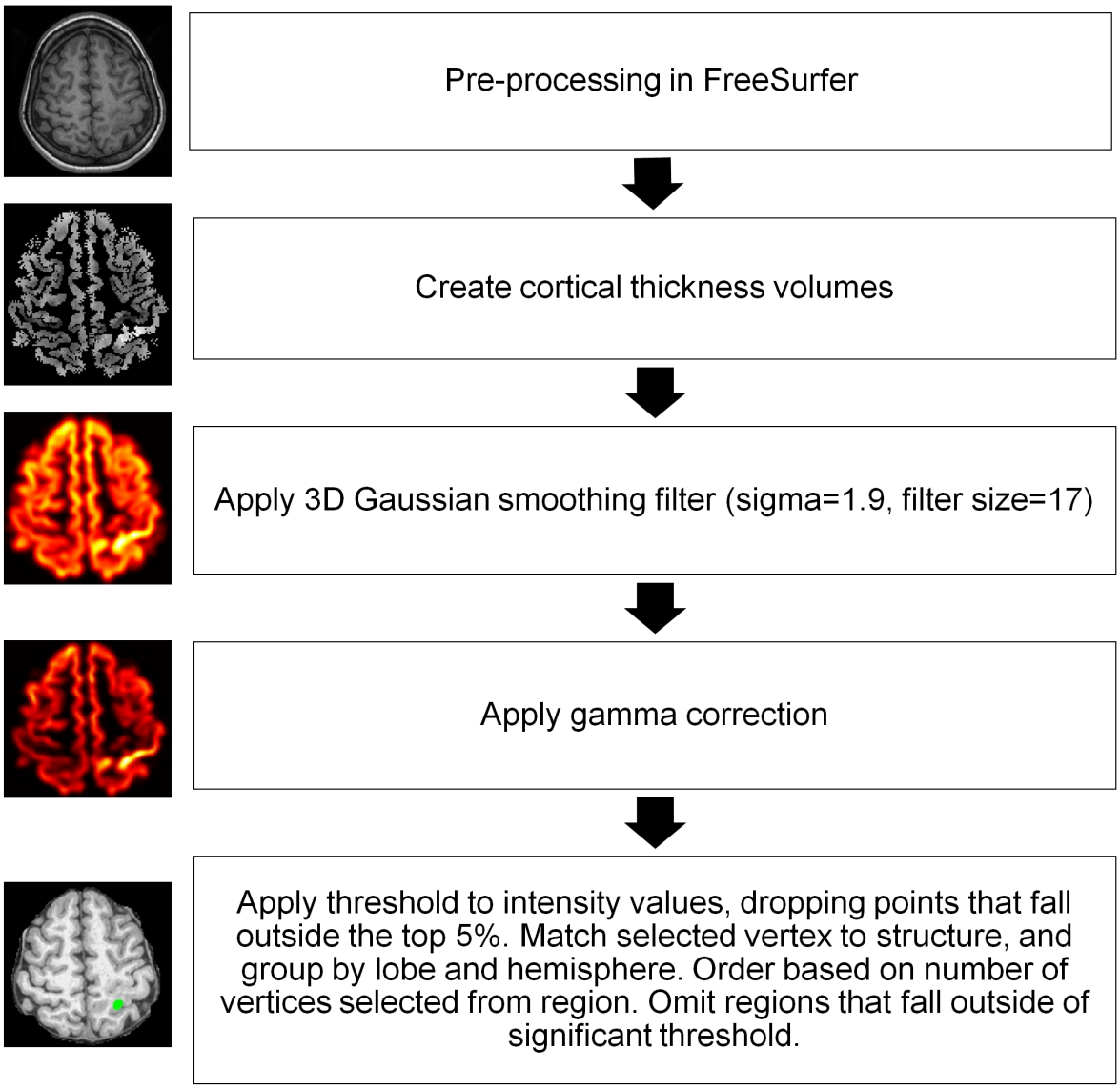


Figure 3.10: Overview of method to identify Epileptogenic lesions using cortical thickness mappings

3.1.2 Identifying Epileptogenic Lesions using Blurring along Gray-White Matter Interface

Blurring at the gray-white matter interface is one of the MRI features of FCD. We seek to verify if using the objective measure of blurring of the GM/WM interface could assist with identifying FCD on T1-weighted MRI scans. To identify the location of the interface the FreeSurfer produced ribbon.mgz volumes was used. The ribbon.mgz volume is a binary mask of the cortical ribbon that maps the left/right hemisphere gray matter and white Matter. To detect the interface, at each voxel the neighbouring 8 voxels were examined. If 30% were contained in the Gray Matter and 30% were contained in the White Matter, the voxel was considered to fall on the interface. A binary mask volume of the interface was composed for each case (Figure 3.11). A sample result of the generated interface mask of a T1-weighted acquisition is presented in Figure 3.12a and 3.12b.

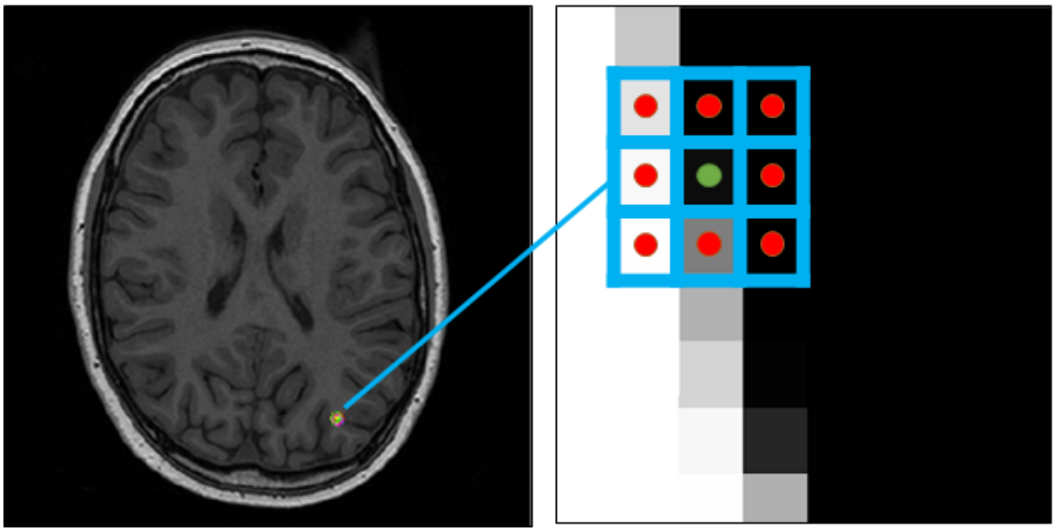
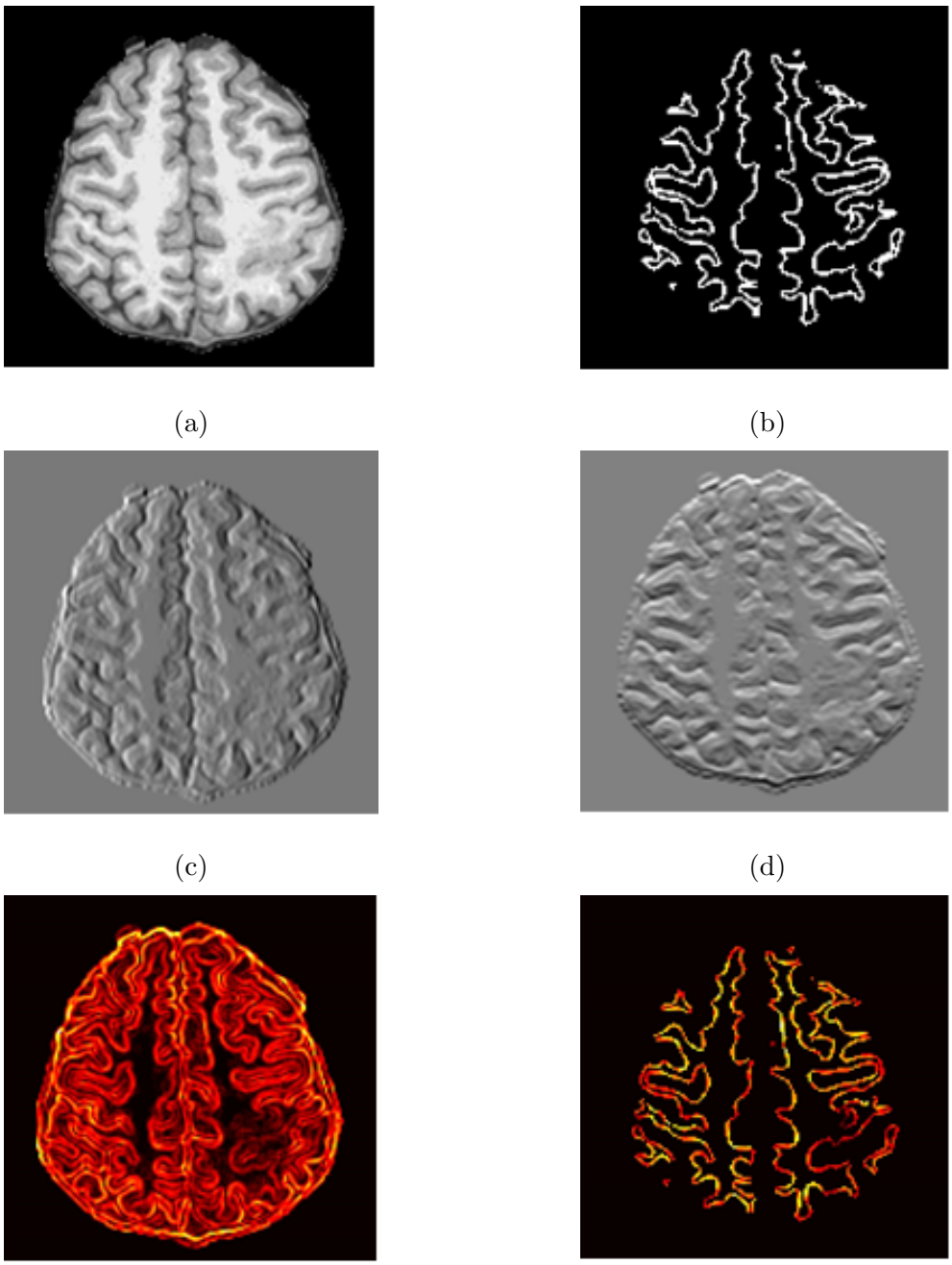


Figure 3.11: Method of selecting values which fall on interface. Consider a region of interest of single axial slice, for a given central voxel (indicated by green dot), the 8 neighbouring voxels (indicated by red dots) are examined to determine if they are contained in the gray matter or white matter.



(a) (b) (c) (d) (e) (f)

Figure 3.12: (a) Original T1-weighted image, (b) Sample interface mask generated, (c) Directional gradient along x-axis (G_x), (d) Directional gradient along y-axis (G_y), (e) Gradient magnitude of complete image, and (f) Interface mask was used to segment regions of the gradient magnitude, producing the gradient magnitude of the interface.

To model the blur we employed the gradient magnitude, a measure of intensity change at each pixel of an image. It can alternatively be described as a measure of steepness of the slope at each pixel, and is commonly used for edge detection. To compute the gradient magnitude for a 2-D image $I(x, y)$ we first compute the x and y directional gradients (refer to Figure 3.12c and 3.12d) by taking the partial derivatives such that

$$I_x = \frac{dI(x,y)}{dx} = \frac{I(x+1,y) - I(x-1,y)}{2} \tag{3.1}$$

and

$$I_y = \frac{dI(x,y)}{dy} = \frac{I(x,y+1) - I(x,y-1)}{2}. \tag{3.2}$$

Following this, gradient magnitude can be computed by $GM(x, y) = \sqrt{I_x^2 + I_y^2}$. This is depicted in Figure 3.12e.

The gradient magnitude was computed for each case using the original volume following intensity normalization from FreeSurfer. The interface binary mask was used to segment the interface regions of the gradient magnitude (Figure 3.12f). Blurring is indicated by regions with lower values of gradient magnitude [ACB+03]. For each axial slice, we consider all non-zero values and identify it's maximum and minimum value. The difference of these values is computed, and threshold of 5% was applied to the measurement and added to the minimum. A mask was used to select all non-zero values below this threshold. This process was repeated for each axial slice of a volume producing a vector of selected vertices. The vertices were matched to their structure name and hemisphere using annotation files produced by FreeSurfer during pre-processing. Following this they were combined using coarse labelling to the 8 major lobes. Any lobe which had total selected vertices below 50 was dropped. In all 31 cases, regions were selected, of which 23 cases corresponded to the true area. The lesional sensitivity and specificity were respectively 72% and 67% and the subjectwise specificity was 74%.

	Coarse Labelling - Overall(%)	Fine Labelling - Overall(%)
Lesional Sensitivity	72	68
Lesional Specificity	67	63
Subjectwise Sensitivity	74	68

Table 3.3: Sensitivity and specificity analysis of results using blurring of gray-white matter interface

The selected locations were then mapped using fine labelling of 70 structures. Of the 31 cases all 31 had areas selected, of which 21 corresponded to the true location. This produces lesional sensitivity of 68%, lesional specificity of 63%, and subjectwise sensitivity of 68%. Since this method directly took the output of the lobe droppings that occurred in coarse labeling into the fine labeling, we can also examine the performance of only the cases that were correctly partially/completely colocalized at the first stage. Thus considering the 23 cases that were correctly identified at coarse labelling, their resulting fine labelling produced lesional sensitivity and specificity of 90% and 60%. The results of coarse and fine labelling can be found in Table 3.3.

Sample results are presented in Figure 3.13. In some cases, solely true locations were selected. For a majority of cases the true location was amongst several locations selected. The complete pipeline of the method followed is presented in Figure 3.14.

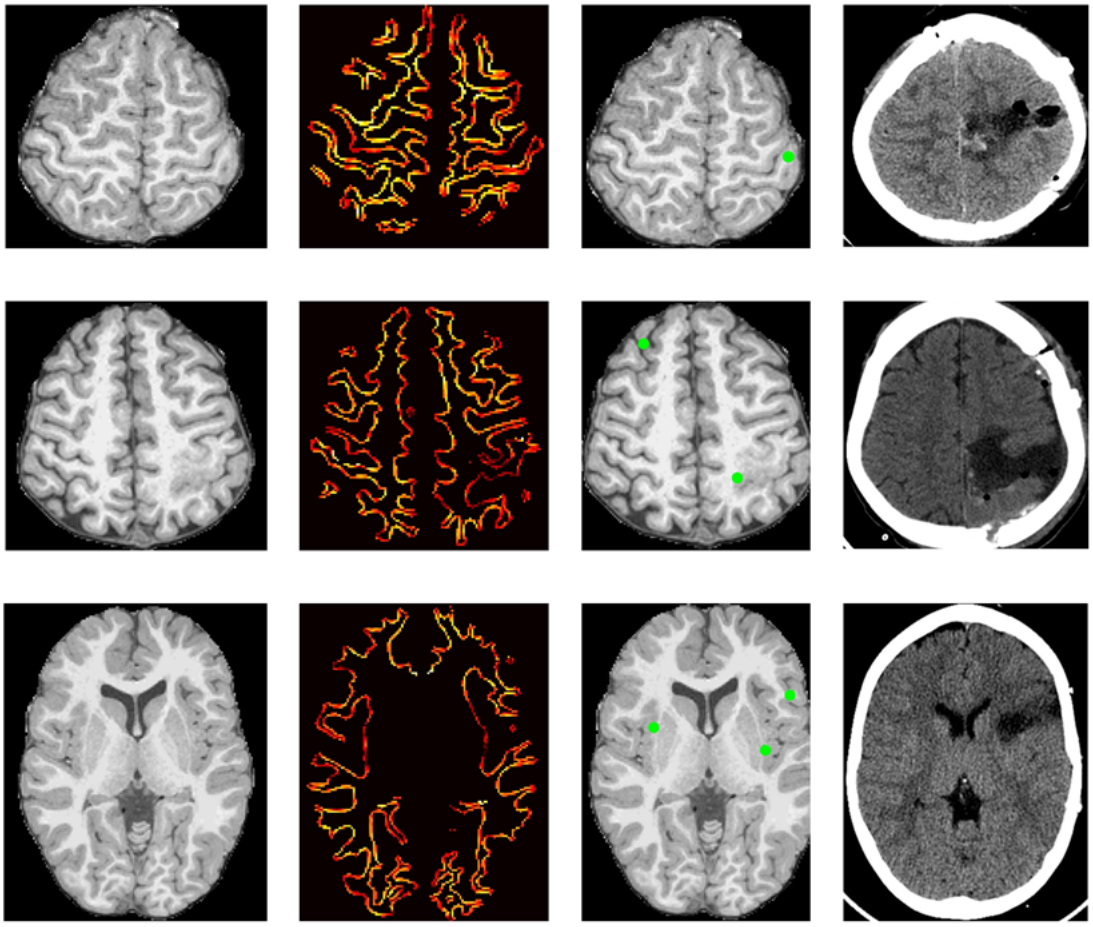


Figure 3.13: Results of volumetric T1 MRI processing using blurring of gray-white matter interface. (Column 1) Original MRI image after normalization and skull stripping , (Column 2) Showing gradient magnitude measurements on interface (Column 3) Original image with green overlay indicating vertices identified by program to indicate lesion, (Column 4) surgical resection site on post-operative CT or T1 scan. Scan indicating location of lesion was removed

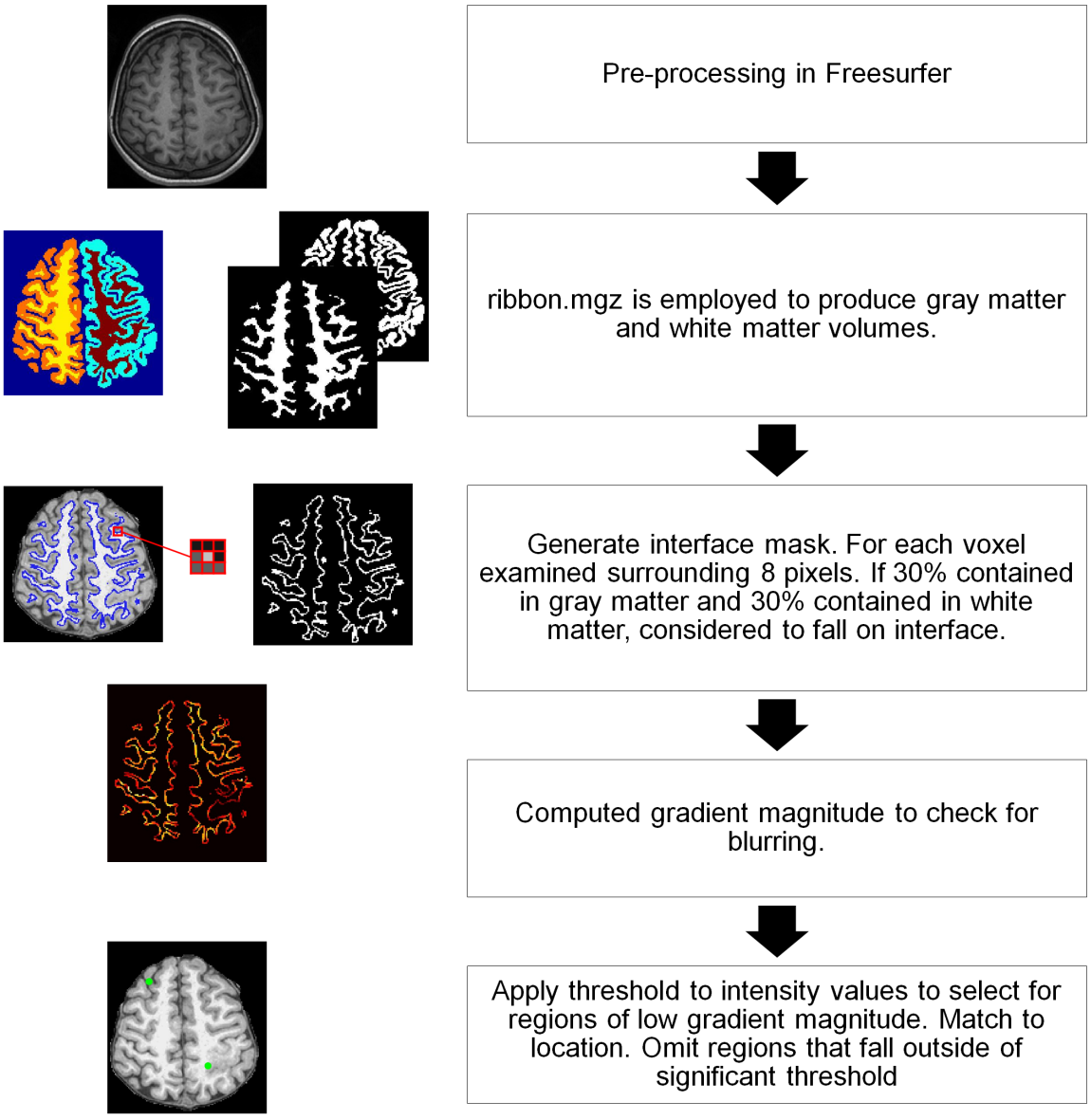
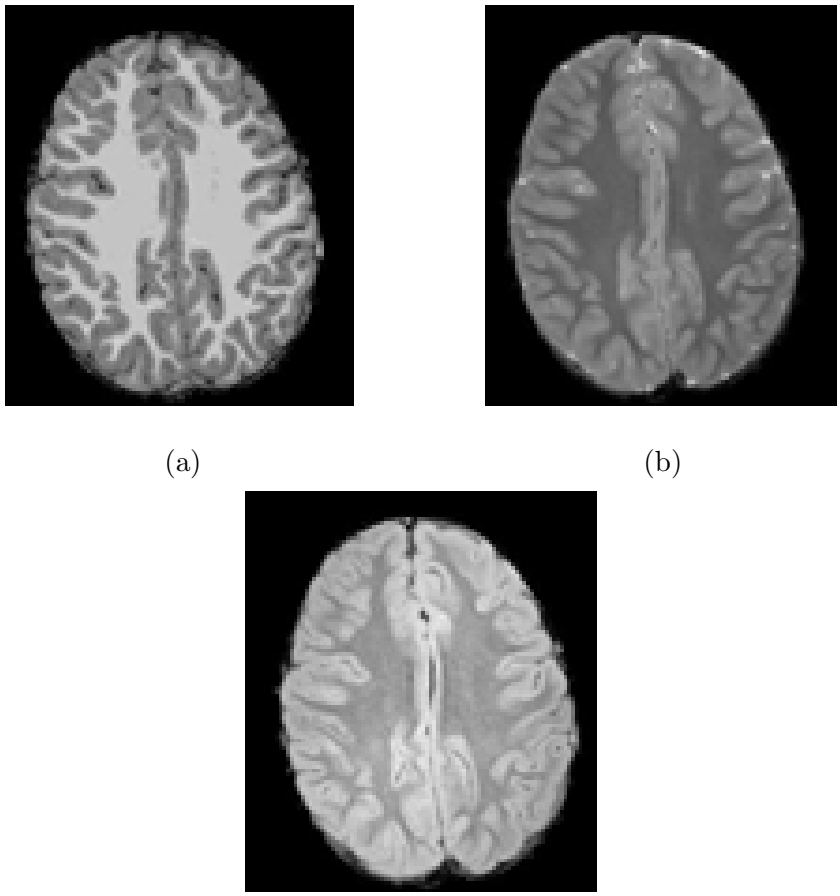


Figure 3.14: Pipeline to identify epileptogenic lesions using blurring along gray-white matter interface

3.1.3 Comparing results of T1, T2, and FLAIR acquisitions using Blurring along Gray-White Matter Interface

Along with the T1-weighted acquisition, other acquisition methods are commonly performed for patients. These modalities can also supply additional information during

diagnosis. For this reason we sought to explore the impact of different modalities at detecting lesion location using gradient magnitude mappings.



(a) (b)

(c)

Figure 3.15: Axial slice of (a) T1-weighted and registered (b) T2-weighted and (c) FLAIR Volumes

The modalities considered were T1, T2, and FLAIR. FreeSurfer can directly process T1 acquisitions, as opposed to T2 and FLAIR. In order to incorporate other supplied modalities into analysis, other means of pre-processing were examined. For T2 and FLAIR acquisitions we performed the following pipeline in preparation of analysis

- In MATLAB, Dicom Import of the SPM package was used to read in the T2 and

FLAIR volumes and save the output as .nii file.

- In the command terminal the FreeSurfer registration tools bbregister and tkregister2 were used with a flag for the subject, to register the .nii to its corresponding T1 acquisition, and output the registration file register.dat.

- Using mri_vol2vol in the command terminal, resampled the skull stripped T1 volume acquired during FreeSurfer pre-processing into the field-of-view of the T2 and the FLAIR volume and saved each as .mgz file.

- Applied the skull stripped mask volume to the T2 and FLAIR using mri_mask and saved .mgz file.

- Used mri_vol2vol to resample the volume to a 256x256x256 space. This will allow for utilizing T1 structural data that we already have from pre-processing in FreeSurfer. Saved as final .mgz volume for use.

Samples results of the axial slices of registered volumes are presented in Figure 3.15.

With the T2 and FLAIR images now in the same space as T1, the structural details from processing the T1 image could be used in combination with all modalities. The gradient magnitude was computed for each case for all 3 modalities and the interface was segmented. Blurring is indicated by regions with lower values of gradient magnitude. For each axial slice, we consider all non-zero values, identify the maximum and the minimum, and compute the difference of the two. In the previous experiment solely considering T1 acquisitions, a threshold of 5% was applied to the measurement and added to the minimum. In order to compare the impact of modality on result, the same method of selection had to be used across the board. One of the issues was that the T2, and FLAIR acquisitions were more sensitive than T1, and using a threshold of 5% was too high. Thus a confidence of 1% was applied to the measurement and added to the minimum of each modality. A mask was used to select all non-zero values below this

threshold. This process was repeated for each axial slice of a volume producing a vector of selected vertices. The vertices were matched their structure name and hemisphere using annotation files produced by FreeSurfer during pre-processing. Following this they were matched using fine labelling to the 70 structural lobes. Any lobe which had total selected vertices below 8 was dropped. The results can be found in Table 3.4. In the case of T1 acquisitions, using this increased sensitivity resulted in clusters identified in 3 patients of the 31, all of which corresponded to surgical resections sites. The lesional sensitivity, lesional specificity, and subjectwise sensitivity were respectively 3%, 99%, and 10%. For the T2 acquisition, clusters were identified in all 31 patients, of which 23 corresponded to surgical resections sites. The lesional sensitivity, lesional specificity, and subjectwise sensitivity were respectively 55%, 80%, and 74%. With the FLAIR acquisition, clusters were identified in all 31 patients, of which 25 corresponded to surgical resections sites. The lesional sensitivity, lesional specificity, and subjectwise sensitivity were respectively 61%, 74%, and 81%. These results imply that FLAIR and T2 better colocalize lesions when using blurring of the interface, and supports the argument for further examination into the use of multiple modalities in computer aided lesion detection techniques.

Modality	Lesional Sensitivity (%)	Lesional Specificity (%)	Subjectwise Sensitivity (%)
T1	3	99	10
T2	55	80	74
FLAIR	61	74	81

Table 3.4: Comparing analysis of T1, T2 and FLAIR

3.2 Identifying Epileptogenic Lesions using Classification Techniques

3.2.1 Background

Preliminary analysis using morphometry with cortical thickness and blurring of the interface produced reasonable results for obvious cases. Volume maps of the features correctly localized lesions in such cases, but performance was poor for instances in which the established features were markedly subtle. Methods combining morphometric and textural features have yielded improved detection rates when compared to standard radiological approaches in the past [ACB+03]. The next step in the approach was to introduce textural features into the analysis and examine if an semi-automated detection model could be developed.

The method described by Antel et al. [ACB+03] utilized a combination of 18 FCD and 14 control cases from an adult cohort. The true location of lesions was manually segmented by an expert to be used in the verification of performance. An automated classifier was developed using a 2-step Bayesian classifier trained on recognized MRI FCD features and textural features derived from GLCMs to classify voxels as lesional or non-lesional. A threshold derived from the largest clusters present in each control case, taken to be the result of random noise, was applied to all subjects (healthy controls and FCD cases) effectively dropping all clusters selected in control cases during the classification process. Any clusters that remained were identified as lesional. The results of this implementation was a subjectwise specificity of 100% and subjectwise sensitivity of 83%. We developed a modified version of this, and implemented a 2-Step Naive Bayes classifier method and a 2-Step SVM classifier method. A comparative study was performed implementing different classification techniques (k-NN, Conditional Inference Trees, Random Forest, and Naïve Bayes) to act as markers for performance with the 2-Step implemen-

tations. These techniques were selected on the basis of implementations performed by others in the field of computer aided epileptogenic lesion selection [BBC+08] [EAHCL13] [HBS+16] [KGL16] [ACB+03]. We will perform a comparative study using preliminary techniques (3.2.3) followed by the implementation of 2-Step methods (3.2.4) and 3.2.5). Results of all methods will be compared in Section 3.2.6. Supplementary experiments examining MRI-negative cases 3.2.7 and the use of other modalities are then performed (3.2.8).

3.2.2 Experimental Setup

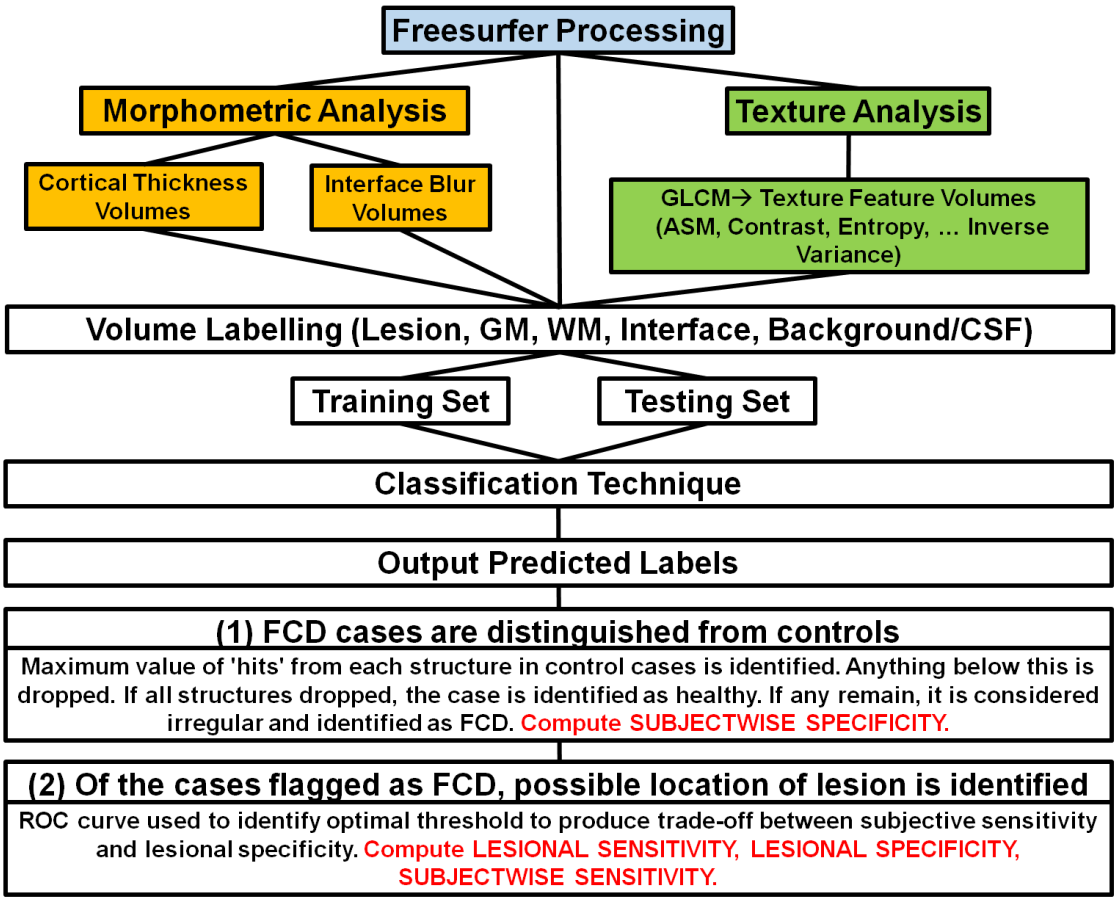


Figure 3.16: Classification Pipeline

We present the complete pipeline to identify lesions in Figure 3.16. The data consisted of 31 MRI-positive FCD cases, and 13 healthy control cases. Quantitatively representing texture was performed using Gray-Level Co-occurrence Matrices (GLCM). Symmetric GLCMs were computed 3-dimensionally on T1-weighted acquisitions with the following parameters: a distance of 3 voxels, and an intensity range of 32 gray levels. The symmetry allowed the 26 directions to be reduced to 13, such that 13 co-occurrence matrices were produced per voxel. From each generated matrix 12 textural descriptor features were computed: Angular Second Momentum, Contrast, Entropy, Correlation, Homogeneity, Variance, Sum Mean, Inertia, Cluster Shade, Cluster Tendency, Max Probability, and Inverse Variance. These 12 descriptor features were the available attributes in *cooc3d*. For each feature, the average of the 13 was taken to be the value mapped back to the position. Figure 3.17 depicts a sample axial slice using the 13 directions to produce the average contrast volume.

Z-score volumes were computed for each textural descriptor attribute average volume by computing $z = \frac{x-\mu}{\sigma}$ where x is the original voxel value, μ the volume-specific mean, and σ the volume specific standard deviation. Once all the volumes for all cases were produced and within-subject z-scores computed, they were normalized by computing between-subject z-scores using the overall mean $\mu_{overall_F}$ and standard deviation $\sigma_{overall_F}$ for each feature $F \in \{F_{ASM}, F_{CONTRAST}, ...F_{INV\_VARIANCE}\}$. This process was performed and data was down-sampled to reduce computation time, with every fifth voxel stored to produce texture volumes. Every fifth voxel was selected as the sampling rate for storage because it was the common factor to be used in later resampling on the basis of tissue class. We employed the computational models for two of the common FCD traits cortical thickness, and blurring of the gray-white matter interface previously described in preliminary experiments to generate volumes. These volumes too were standardized by computing z-scores. For classification we considered the z-score volumes of the established morphometric MRI features and the textural features.

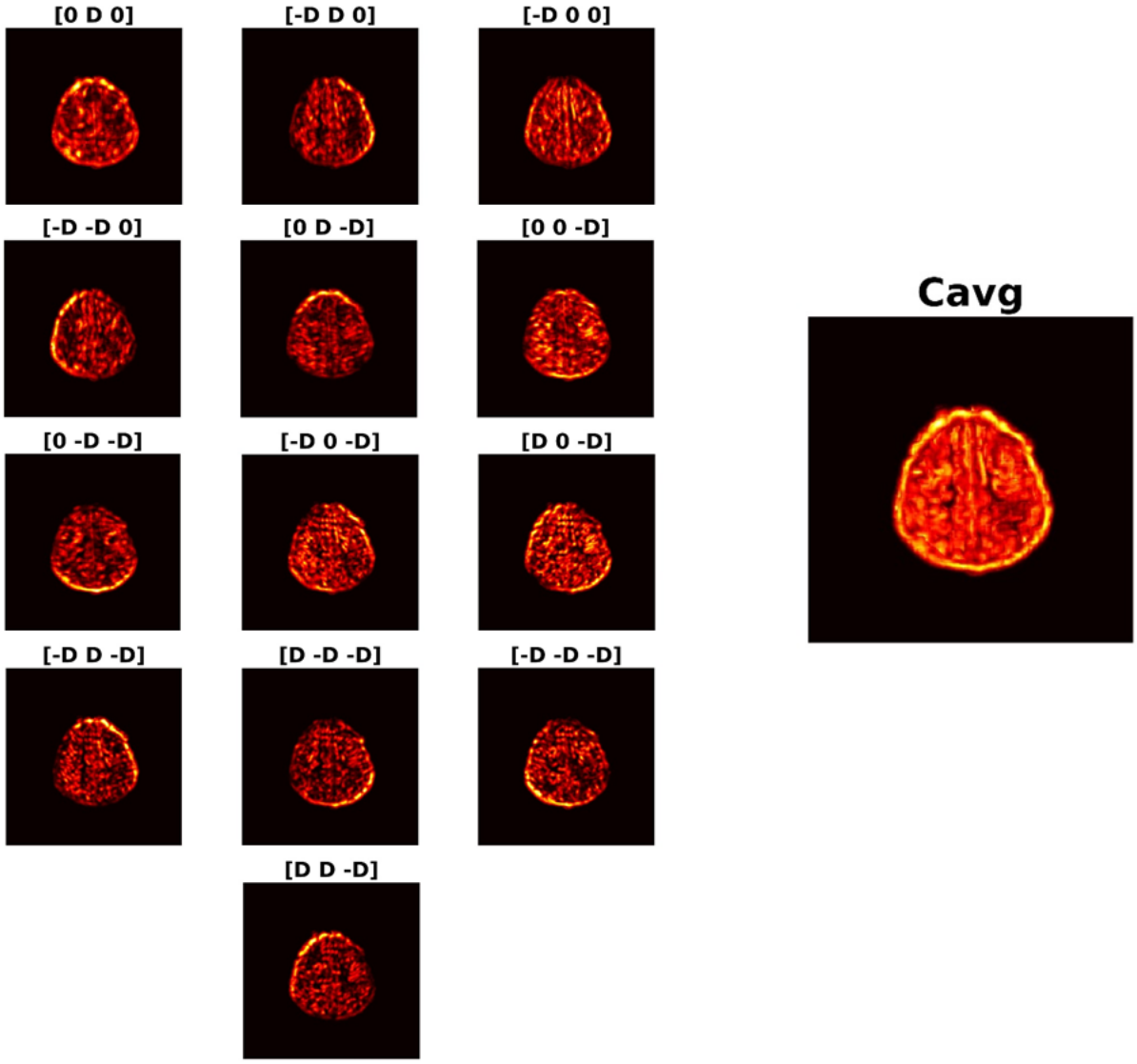


Figure 3.17: Contrast volumes using $D = 3$ for 13 directions and resulting C_{avg}. To generate symmetric 3-D GLCMs, 13 directions are considered. For each direction a matrix is generated and the Contrast is computed. This value gets mapped back to the original voxel around which the GLCMs were generated. The average of the 13 values produced considering each direction gets mapped to the same location in C_{avg}. Doing so for all points produces the average Contrast volume. A sample axial slice of each direction and the generated average for Contrast is displayed.

A voxel in a volume could be labelled by five possible tags: Lesion, Gray Matter, White Matter, Gray-White Matter Interface and Background or Cerebrospinal Fluid (CSF). Using the FreeSurfer generated volumes T1.mgz, ribbon.mgz, and aparc+aseg.mgz in combination with the interface volume previously produced during Blurring along Gray-White Matter Interface experiments, labels could be produced in MATLAB for the Gray Matter, White Matter, Gray-White Matter Interface and Background/CSF (Figure 3.18).

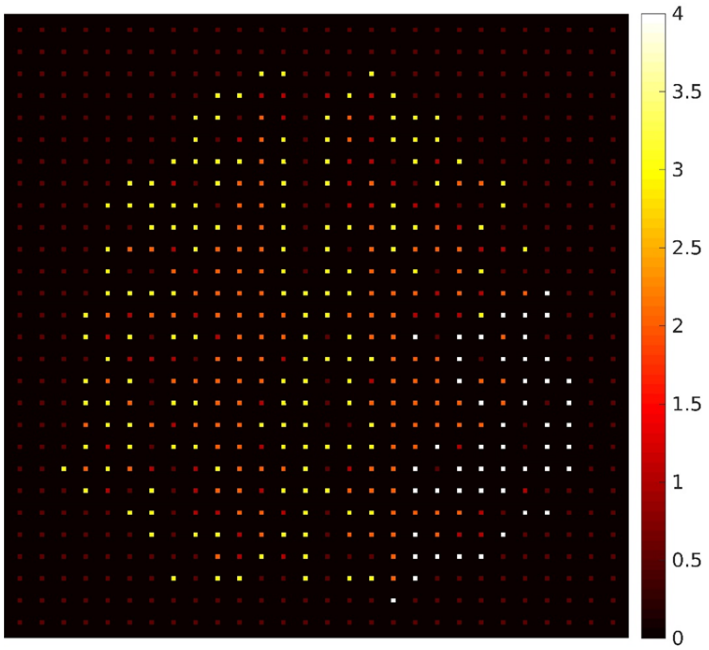


Figure 3.18: Volume labelling. Volumes down-sampled to store every 5th voxel. Displaying sample axial slice with colormap bar of color labels. Labels indicated for Background/CSF (0.5), Gray-white matter interface (1), Cerebral white matter (2), Cerebral cortex gray matter (3), and Lesion (4).

Excel files were created containing subject based lesional information supplied by the radiologist that could be loaded into MATLAB. The Gray Matter, White Matter, Gray-White Matter Interface and Lesion are considered to make up the brain's volume. Two datasets were constructed for classification, a training and testing set, with sampling

rates differing between the two. We seek to train the classifier on a training set balanced with respect to label classes, and test performance on a testing set unbalanced in this respect. Sampling frequencies of the training set data are selected to achieve roughly equal numbers of voxels in each label. For the training set, voxels are uniformly selected from all label classes. The training set consisted of irregular sampling from each of the labelling types such that: every 5th voxel of the lesion was selected, every 15th voxel of the Gray Matter, every 15th voxel of the White Matter, every 5th voxel of the Gray Matter-White Matter Interface, and every 500th voxel of the Background/CSF resulting in approximately equal selection from each label. The testing set utilized equal sampling of the brain volume by selecting every 5th voxel of the brain's complete volume. In addition to this, every 100th voxel of the background (not including CSF) was selected. The sampling frequencies were used to select the corresponding observed attribute values at a given voxel location. Each dataset (training and testing) consisted of 20 columns: Case, FreeSurfer Structure Identifier, Class Label (1 and 0 referencing lesional and non-lesional respectively), Coronal Vertex, Axial Vertex, Sagittal Vertex, and the 14 features.

Classification was performed using either a standard single step method using all features as input, or a modified 2-Step method following the approach described in [ACB$^+$03]. In the 2-step method, data is first trained solely on morphometric features. Data that is classified as lesional is then reclassified using textural features. For all classification techniques, we can generalize the approach as follows: Given a training and testing data set, we classify the test data in R as lesional or non-lesional using a leave-one-out $(N-1)$ approach for each subject $N \in \{1, 2, 3, ...44\}$ (31 MRI-Positive Case, 13 Healthy Controls). For each experiment the process was repeated 5 times to measure computation time. Results of experiments for each classification technique are written to .csv files containing voxel information and predicted class labels to be read into MATLAB. These results are used to implement a 2-Step pipeline in which cases are first identified as either control or FCD, and then the structural location of the lesion

is identified. To do so, control cases are examined for the maximum cluster size within each of the structures to introduce structural thresholds. Lesional clusters smaller than these thresholds within each structure are deemed a result of random noise and omitted from the lesion map produced by the classifier. Thus, for each case, the thresholds are applied to drop structures. If all structures remain unflagged, the case is identified as a control. If any region is flagged, the case is deemed irregular and identified as an FCD case. For all cases identified as FCD, the next step is to find the possible location of the lesion. To do this we select an appropriate threshold using criterion selection via ROC curves. We previously discussed the Youden index (J) as a means of selecting the optimal parameter of a ROC curve [Eke12]. We can express J relative to subjectwise sensitivity and subjectwise specificity in terms of the $(TP)_i$, $(TN)_i$, $(FP)_i$, and $(FN)_i$ relating to the i-th subject as

$$\begin{aligned} J &= \max\left((\text{Subjectwise Sensitivity})_i + (\text{Subjectwise Specificity})_i - 1\right) \\ &= \max\left(\frac{\sum_{i=1}^{N}[(TP)_i > 0]}{N} + \frac{\sum_{i=1}^{N}[(TN)_i > 0]}{N} - 1\right). \end{aligned} \tag{3.3}$$

Due to the nature of our pipeline for classification, subjectwise sensitivity and subjectwise specificity are computed in isolated steps (Refer to Figure 3.16). A control and FCD case are distinguished and subjectwise specificity is computed. In the next step, FCD cases have structures dropped using an optimal threshold, and subjectwise sensitivity, lesional sensitivity and lesional specificity are computed. Once a method produces an appropriate level of subjectwise specificity we seek to produce a strong subjectwise sensitivity, however some measure of specificity must still be incorporated into it. Thus we employ a modified Youden's J index (Eq. 3.4) that aims to produce a balance between subjectwise sensitivity and lesional specificity.

$$\begin{aligned} J_{\text{mod}} &= \max\left((\text{Subjectwise Sensitivity})_i + (\text{Lesional Specificity})_i - 1\right) \\ &= \max\left(\frac{\sum_{i=1}^{N}[(TP)_i > 0]}{N} + \frac{\sum_{i=1}^{N}(TN)_i}{\sum_{i=1}^{N}[(TN)_i + (FP)_i]} - 1\right) \end{aligned} \tag{3.4}$$

To select the optimal threshold for each algorithm implementation we utilized this modified J index. For all cases identified as FCD, we set the preferred threshold as determined by the ROC curves and a modified J index. Any regions below this are dropped.

3.2.3 Preliminary Performance Assessment Experiment - Comparative Study

We seek to compare the performance of several methods of classification for preliminary assessment and serve as markers for our 2-Step implementations described in subsequent subsections. To perform lesion identification these classification techniques were implemented in R. The methods were k-NN, Conditional Inference Tree, Random Forests, and Naïve Bayes. A leave-one-out cross validation method on the basis of subjects was performed, where each subject N in the test set was classified using a classifier trained on the remaining $N-1$ subjects. This method was employed in all classification systems. The k-Nearest Neighbour algorithm was implemented utilizing the package *class*. The package *partykit* was used to employ a Conditional Inference tree for classification. A Random Forest algorithm was performed with the package *randomForest*. To execute Naïve Bayes the package *naivebayes* was used. All algorithms were trained on all 14 features.

The effect of parameters on performance of k-NN and Random forest was inspected. We previously discussed the selection of an optimal k value for k-NN being a key consideration, especially in instances of imbalanced problems (as is the case with lesion detection). We seek to balance the effect of noise, without the classifier being biased towards prediction of the majority non-lesional class. With the Random Forest algorithm we seek to identify the number of trees that offers predictive power without unnecessary computation time. Processing time is also a factor with the k-NN algorithm, with larger k values naturally leading to longer runtimes.

To examine the impact of k, k-NN was performed for $k = 2^j$ for $j = 1, 2, \ldots 8$. The

impact of the number of trees, t, on results was similarly performed for Random Forest with $t = 2^j$ for $j = 1, 2, \ldots 8$. The runtime results are presented in Figure 3.19 and Figure 3.20. The associated lesional sensitivity, lesional specificity, subjectwise sensitivity and modified Youden's index at each instance of evaluation for each algorithm is presented in Figure 3.21 and Figure 3.22. Seeking a balance between processing time and the modified J index of the ROC, resulted in the selection of $k = 16$ for k-NN and $t = 16$ for Random Forest for our purposes. These parameter values are utilized in the experiments for comparative analysis.

	k-NN	Conditional Inference Tree	Random Forest	Naive Bayes
Subjectwise Specificity (%)	100	100	100	100
AUC	0.87	0.86	0.85	0.88
Selected Threshold	40	7	26	30
Lesional Sensitivity (%)	67	51	81	52
Lesional Specificity (%)	73	79	67	78
Subjectwise Sensitivity (%)	94	81	94	84

Table 3.5: Evaluating performance of k-NN, Conditional Inference Trees, Random Forest, and Naive Bayes classification

Four experiments were implemented with the different classification techniques (1) k-NN with $k = 16$, (2) Conditional Inference Tree, (3) Random Forests with number of trees=16, and (4) Naïve Bayes without any prior values. The pipeline described in Figure 3.16 was followed. ROC curves with selected modified J indices are presented in Figure 3.24. Runtime analysis is presented in Figure 3.23. Results are presented in Table 3.5 and indicate that all methods perform comparably. The healthy controls were all correctly identified, and substantially high AUC values were produced by each method. The use of k-NN produced the highest AUC value, indicating the most accuracy. The implementation using random forest had the highest subjectwise sensitivity and lesional sensitivity indicating that it was able to most completely colocalize the most lesions. However it also had the lowest lesional specificity, indicating that improved detection is at the cost of precision. The highest lesional specificity was seen with the use of conditional inference trees. Although it also had the lowest subjectwise sensitivity overall, it is still a relatively high value (81%). While it has low lesional sensitivity, the conditional inference tree method is able to more accurately identify lesional structures. The Naive Bayes implementation performs relatively on par with the conditional inference tree method.

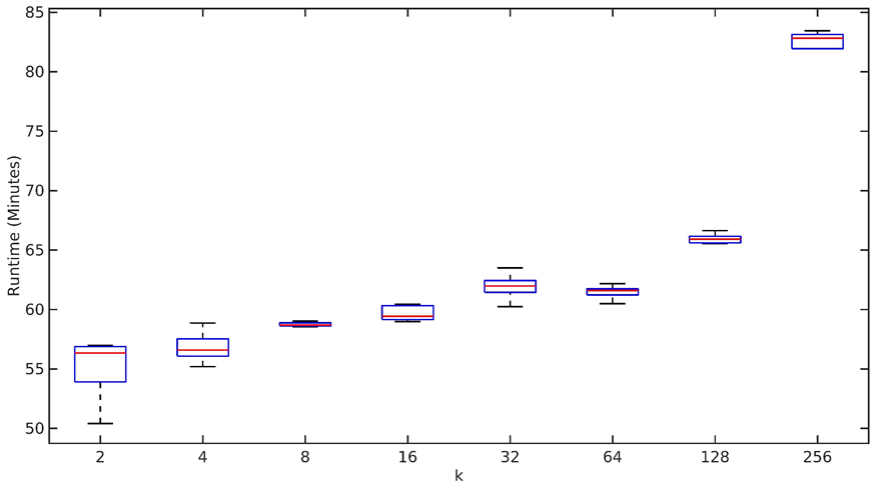


Figure 3.19: Evaluating k-NN runtime with respect to parameter selection. The k-NN algorithm was implemented with k increasing exponentially using base 2 for $k = 2^j$ for $j = 1, 2, \ldots 8$. This process was repeated 5 times, and results are presented in boxplots.

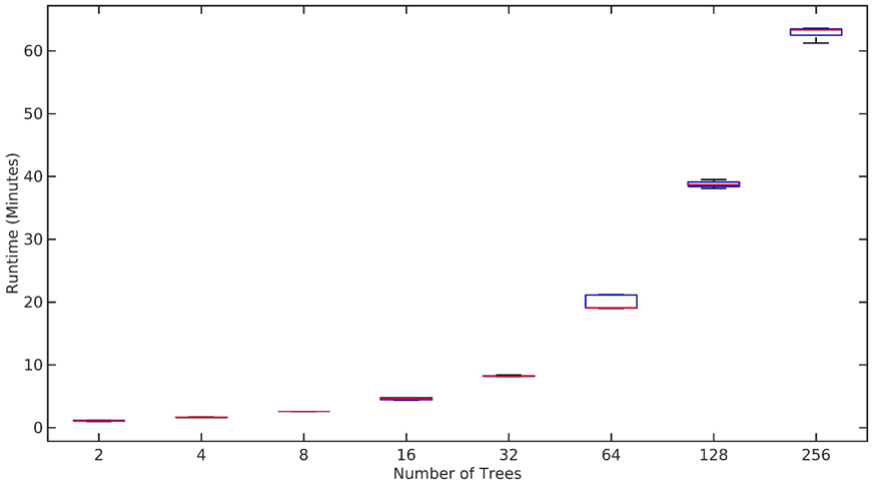


Figure 3.20: Evaluating Random Forest runtime with respect to parameter selection. The Random Forest algorithm was implemented with t increasing exponentially using base 2 for $t = 2^j$ for $j = 1, 2, \ldots 8$. This process was repeated 5 times, and results are presented in boxplots.

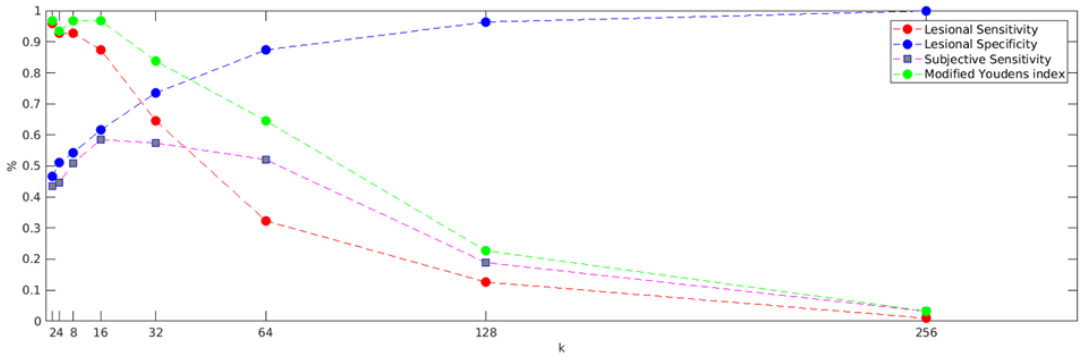


Figure 3.21: k-NN performance analysis. Generated multiple random forests with number of trees k increasing exponentially using base 2 for $k = 2^j$ for $j = 1, 2, \ldots 8$. The associated lesional sensitivity, lesional specificity, subjectwise sensitivity and modified youden's index is presented. We seek to maximize the modified Youden's index (which seeks a balance between the subjectwise sensitivity and lesional specificity) for an appropriate computation time (Figure 3.19). This occurs at $k = 2^4 = 16$.

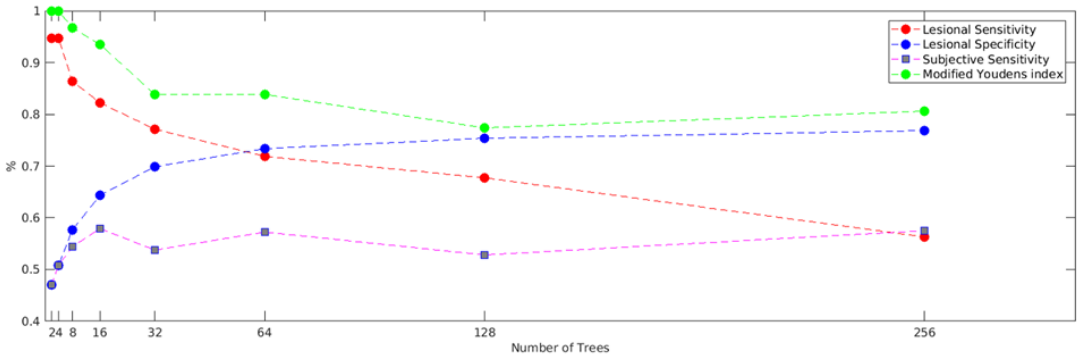


Figure 3.22: Random Forest performance analysis. Generated multiple random forests with number of trees t increasing exponentially using base 2 for $t = 2^j$ for $j = 1, 2, \ldots 8$. The associated lesional sensitivity, lesional specificity, subjectwise sensitivity and modified Youden's index is presented. We seek to maximize the modified Youden's index (which seeks a balance between the subjectwise sensitivity and lesional specificity) for an appropriate computation time (Figure 3.20). This occurs at $t = 2^4 = 16$.

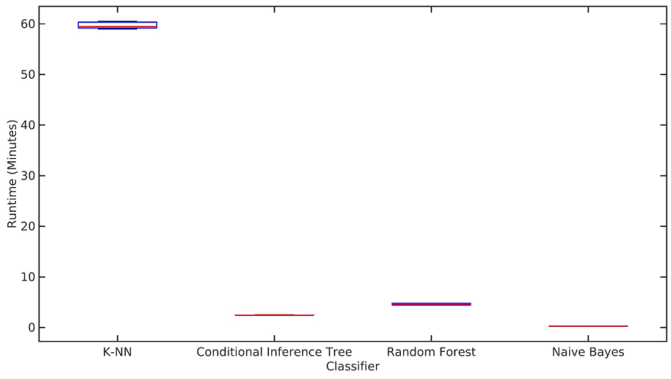


Figure 3.23: k-NN with $k = 16$, Conditional Inference Tree, Random Forest with $t = 16$ and Naive Bayes Classifier runtime

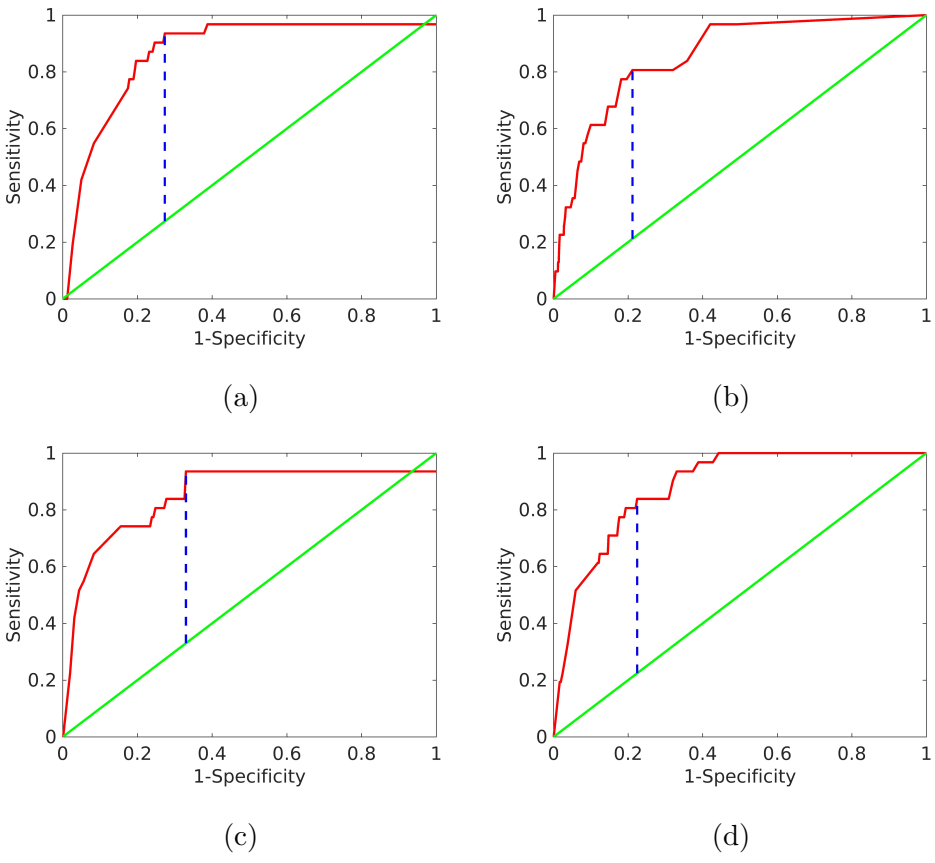


Figure 3.24: (a) k-NN, (b) Conditional Inference Tree, (c) Random Forest and (d) Naive Bayes Classifier ROC curves with threshold selection using modified J index

3.2.4 2-Step Naive Bayes Classifier

	2-Step Naive Bayes
Subjectwise Specificity (%)	100
AUC	0.90
Selected Threshold	17
Lesional Sensitivity (%)	63
Lesional Specificity (%)	75
Subjectwise Sensitivity (%)	94

Table 3.6: 2-Step Naive Bayes Classifier

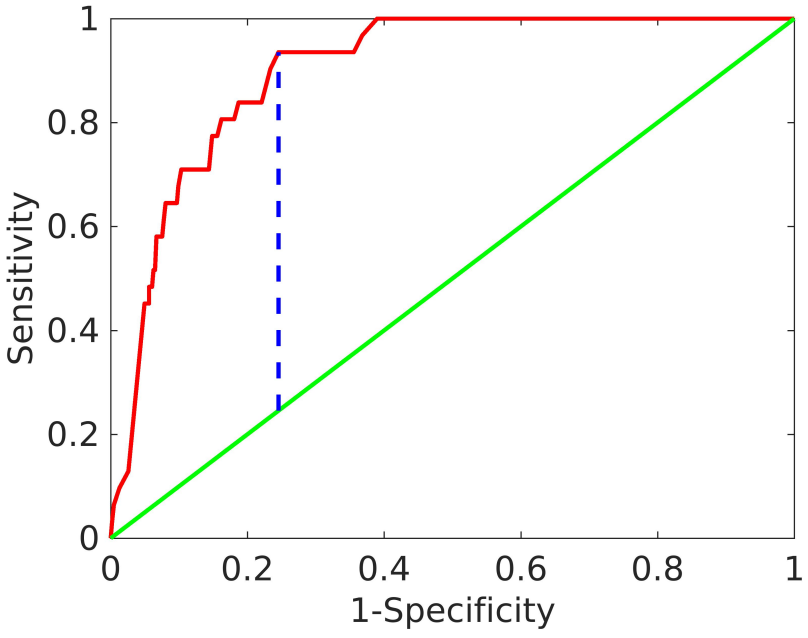


Figure 3.25: 2-Step Naive Bayes Classifier ROC curve with threshold selection using modified J index

Antel et al. [ACB+03] implemented a 2-Step Bayesian classifier in which attribute data was split into morphometric and textural features, and the classifier is first trained on morphometry then using features from GLCMs. A two-classifier approach is justified because the two groups of attributes extract different types of information from the MRI volume. The computational models of MRI morphometric characteristics of FCD (i.e. cortical thickness and blurring of the gray-white matter interface) measure first-order statistical properties of individual voxels. The texture features measure second-order properties of the volume by assessing spatial relationships amongst voxels of varying intensities using the associated GLCMs. Therefore, the use of two classifiers is warranted, as they would be highly trained in distinct regions within the feature space. Their method produced high values of subjectwise sensitivity (colocalized lesion in 15/18 of the FCD subjects) and subjectwise specificity (rejected all Control subjects). For this reason, the paper served as a key motivation in our experimental setup. We implemented a modified version of this method to perform classification using Naïve Bayes to create a 2-Step classifier. The classifier is first trained on the morphometric features with priors were set to heighten lesional sensitivity to encourage selection. This was associated with a cost of some misclassification with Gray Matter. Data classified as lesional by morphometry is then reclassified by a secondary classifier trained solely on the second-order texture features. ROC curves with selected modified J indices are presented in 3.25. The results of the experiment are displayed in Table 3.6.

3.2.5 2-Step SVM Classifier

	2-Step SVM
Subjectwise Specificity (%)	100
AUC	0.88
Selected Threshold (%)	54
Lesional Sensitivity (%)	69
Lesional Specificity (%)	72
Subjectwise Sensitivity (%)	94

Table 3.7: 2-Step Support Vector Machine Classifier

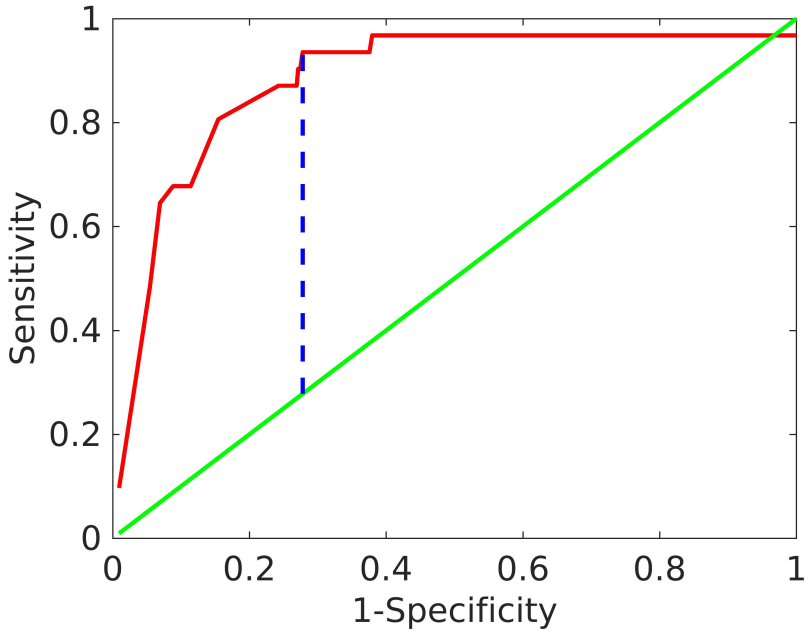


Figure 3.26: 2-Step SVM Classifier ROC curve with threshold selection using modified J index

Support Vector Machines (SVM) have been used in identifying epileptogenic lesions and performed well in MRI-negative cases. They were found to be more specific in region localization with fewer false positive detections generated when compared to SPM implementations [EAHJ$^+$16] [EAHCL13] [HBS$^+$16]. The performance of the 2-Step Naive Bayes Classifier met expectations, thus the idea of applying the concept of a 2-Step classifier using SVM was explored. [HBS$^+$16] employed 2 separate SVM classifiers albeit with a searchlight-based multivoxel pattern analysis methodology using mean thickness and curvature z-scores. The SVM technique was initially not implemented as attempts to do so proved computationally expensive and made it inefficient for the complete data set. However by following the 2-Step method with downsampled data, the notion was that it would be faster to implement as the initial training on the entire data set would be using only 2 attributes, and the secondary training stage would have a marked reduction in the data even though overall, 14 attributes would be used. To implement SVM the *kernlab* package was used with a Gaussian kernel with $\sigma = 3$. ROC curves with selected modified J indices are presented in 3.26. Results are presented in table 3.7.

3.2.6 Evaluating Technique Performances

Table 3.8 present performance measures and runtimes of implemented methods. From the experimental setup previously presented in Figure 3.16, following classification, a 2-step pipeline is applied to the output. In step 1, the focus is on discriminating between Controls and FCD cases. In Table 3.8, we see that all implementations will correctly identify Controls (subjectwise specificity = 100%), however only 3 of the techniques (Conditional Inference Trees, Naive Bayes, 2-Step Naive Bayes) correctly identified all FCD cases. In step 2, a threshold is selected using ROC curves and the maximum modified J index to produce a balance in maximizing subjectwise sensitivity and lesional specificity. We see here that the maximum subjectwise sensitivity is 94 % occurs for 4 of the techniques (k-NN, Random Forest, 2-Step Naive Bayes, 2-Step SVM) implying that in 29/31 FCD

	k-NN	Conditional Inference Tree	Random Forest	Naive Bayes	2-Step Naive Bayes	2-Step SVM
Mean Runtime (minutes)	60	3	5	0.3	0.2	17354
	STEP 1					
Subjectwise Specificity (%)	100	100	100	100	100	100
Number of Incorrectly Rejected Cases of FCD	1	0	2	0	0	1
	STEP 2					
Threshold Selected Using ROC	40	7	26	30	17	54
Modified J index	0.66	0.60	0.61	0.61	0.69	0.66
Lesional Sensitivity (%)	67	51	81	52	63	69
Lesional Specificity (%)	73	79	67	78	75	72
Subjectwise Sensitivity (%)	94	81	94	84	94	94

Table 3.8: Evaluating technique performances between classification techniques. In step 1 input subjects are distinguished between cases healthy controls and FCD subjects. From this step we compute subjectwise specificity and make note of any FCD cases incorrectly identified as healthy subjects. In step 2 using cases identified as FCD in combination with ROC curves of lesional specificity and subjectwise sensitivity a modified J index is computed to select the optimal threshold. The associated lesional sensitivity, lesional specificity and subjectwise sensitivity for this cut-off are noted. The classification techniques were performed 5 times and the mean was noted.

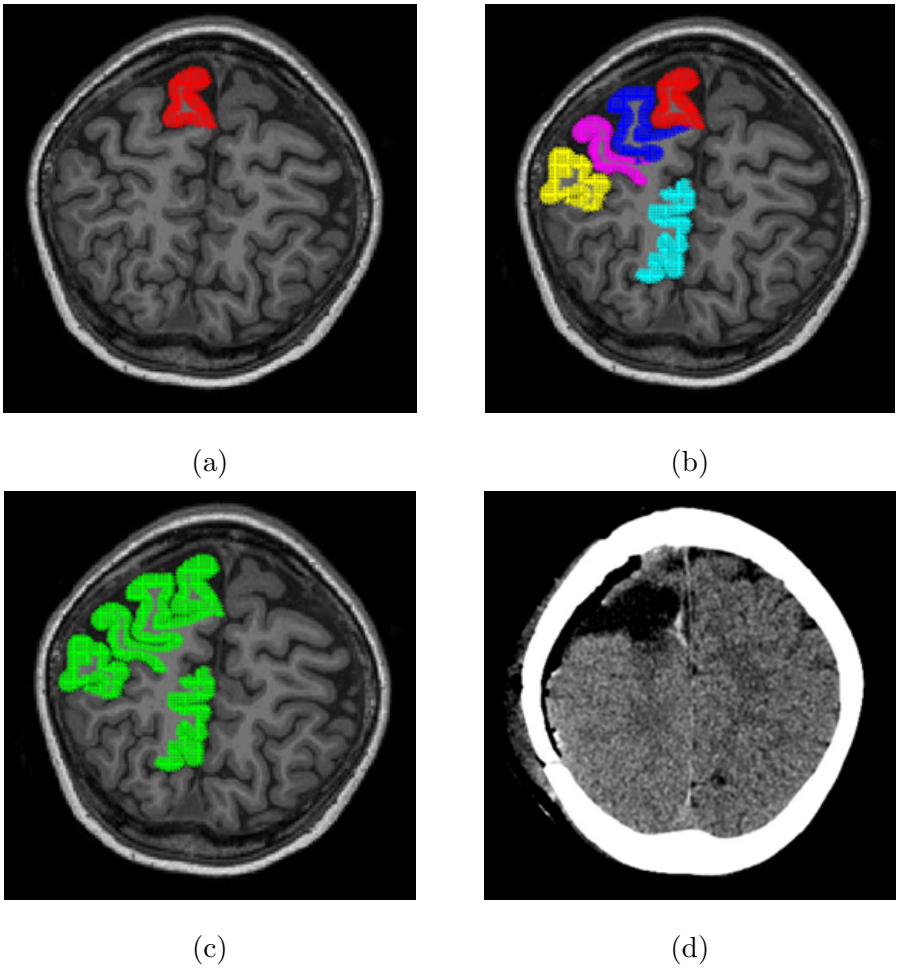
cases the lesions were partially or wholly colocalized. Examining the associated Lesional Specificity for these techniques, k-NN, Random Forest, 2-Step Naive Bayes, 2-Step SVM respectively had 73%, 67%, 75%, and 72%. Thus, the 2-Step Naive Bayes technique produce the best results comparatively, with the maximum subjectwise sensitivity and lesional specificity (reflected by the maximum Modified J Index value). It also did not reject any FCD cases in the first step, and has the shortest mean runtime due to the package being a high performance implementation (0.2 minutes). It should be noted that the results indicate that the SVM underperforms compared to other techniques. However the results of SVM were based on preliminary parameter setting, and are not optimized by any means. Overall, all techniques performed relatively on par with on another, with the 2-Step Naive Bayes producing the largest subjectwise specificity, subjectwise sensitivity, and lesional specificity concurrently. The largest lesional sensitivity was produced by the Random forest implementation, but it was also coupled with the smallest lesional specificity.

3.2.7 A Study of MRI-negative Cases using 2-Step Naive Bayes Classifier

Thus far, we have performed analysis on MRI-positive cases. For MRI-positive cases we noted in preliminary morphometric experiments that established features that were not used to identify the lesion using standard radiological assessment could be detected through computational methods. For this reason we seek to extend the methods used on MRI-positive cases to the increasingly subtle MRI-negative cases. We are satisfied with the pipeline built off MRI-positive cases and controls using the 2-Step Naïve Bayes classifier, and now seek to conduct tests on MRI negative Cases. Four MRI Negative cases (NL041, Nl042, NL043, NL044) were classified using the 2-Step Naïve Bayes classifier following the pipeline presented in 3.16.

The method correctly identified NL043 and NL044 as cases of FCD and selected le-

sional structures colocalizing the true lesion locations in both cases. However, the method incorrectly identified NL041 and NL042 as healthy cases. Thus solely considering the 4 MRI-Negative cases there was a 50% subjectwise sensitivity. Evaluating performance considering solely NL043 and NL044, the lesional sensitivity was 60% and the lesional specificity was 87%. Considering overall performance (accounting for all 4 cases) these values become 25% and 94%. The reason for the drop in lesional sensitivity is due to the face that NL041, NL042, NL043, NL044 respectively had 6, 1, 1, and 4 true lesion structures. Thus failing to recognize NL041 and NL042 contributes more to the lesional sensitivity. The implication of this test is that for the cases that are not rejected (NL043, NL044) the method performs acceptably in terms of colocalizing the lesion while maintaining an appropriate specificity. Sample results are presented in Figure 3.27 for case NL043.



(a) (b)

(c) (d)

Figure 3.27: Results of the true lesion structural location of NL043 and the selected structures. (a) Displays select T1 axial slice with corresponding FreeSurfer region (rightsuperiorfrontal) of lesion highlighted in red. (b) Select T1 axial slice with selected FreeSurfer regions using 2-Step Naive Bayes classification method with colours differing on the basis of region selected (magenta - rightpostcentral, blue - rightprecentral, cyan - rightpre-cuneus, red - rightsuperiorfrontal, Yellow - rightsupramarginal). (c) Select T1 axial slice with selected FreeSurfer regions using 2-Step Naive Bayes Classification method with all selected structures in green. (d) Associated post-surgical CT scan displaying resected location.

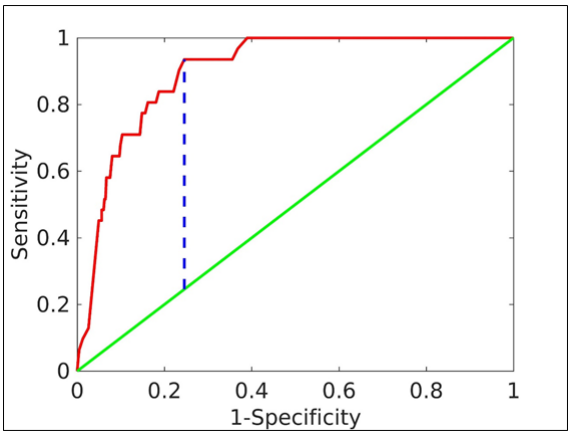
3.2.8 Comparing Performance of 2-Step Naive Bayes Classifier for T1, T2, FLAIR implementations

	T1	T2	FLAIR
	STEP 1		
Subjectwise Specificity (%)	100	100	100
Number of Incorrectly Rejected Cases of FCD	0	1	6
	STEP 2		
Threshold Selected Using ROC	17	22	57
Modified J index	0.69	0.70	0.60
Lesional Sensitivity (%)	63	60	42
Lesional Specificity (%)	75	80	89
Subjectwise Sensitivity (%)	94	90	71

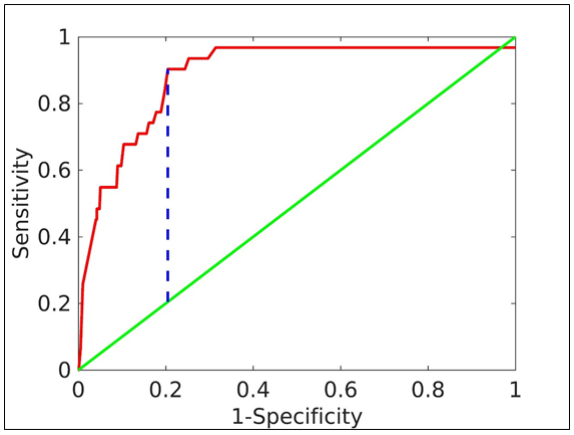
Table 3.9: Evaluating 2-Step Naive Bayes performances using T1, T2, FLAIR

The abnormal signal in cortex may not be well detected on T1-weighted when compared to T2-weighted acquisition or FLAIR. T2-weighted sequences are most sensitive to brain pathology detection. Textural and morphological analysis was performed on volumetric T2, and FLAIR images using the 2-Step Naive Bayes classifier with the same method described in Figure 3.16. Although volumetric T1, T2 and FLAIR sequences were avail-

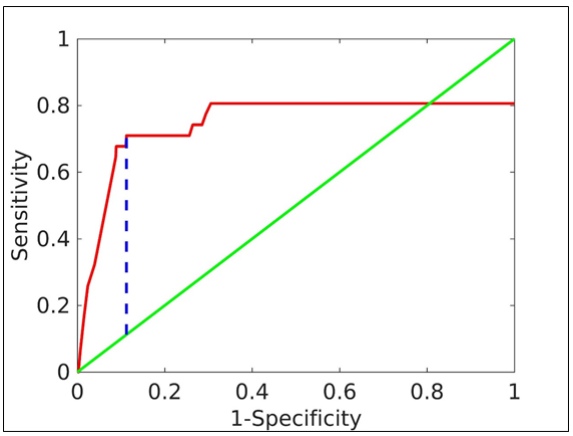
able for all FCD subjects, this was not the case for the control data. While T2-weighted images were available for the same previously used T1 controls, FLAIR sequences were not available. Thus a supplementary set of 13 FLAIR healthy control subjects was used. Note that cortical thickness volumes produced by FreeSurfer was used for all 3 methods, however interface blurring gradient magnitude and texture attribute volumes were derived using the considered modality. ROC curves with selected modified J indices are presented in 3.28. Results are presented in Table 3.9. The respective subjectwise sensitivity was 94 % (29/31 cases colocalized), 90 % (28/31 cases colocalized), and 71% (22/31 cases colocalized) using the T1-weighted, T2-weighted and FLAIR images. The cases in which the T1 method failed to recognize the region were NL011 and NL024. T2 failed in cases NL011, NL024, and NL025. FLAIR failed in cases NL001, NL002, NL008, NL012, Nl013, NL16, NL023, NL024 and NL035.



(a)



(b)



(c)

Figure 3.28: Comparing 2-Step Naive Bayes Classifier ROC curves with threshold selection using modified J index for (a) T1-weighted, (b) T2-weighted, and (c)FLAIR

Chapter 4

Conclusion and Future Work

4.1 Conclusion

We first explored preliminary morphometry experiments (Section 3.1) on solely MRI-positive cases of FCD. In Section 3.1.1 we presented identifying epileptogenic lesions using computational models of cortical thickness with thresholding. The resulting subjectwise sensitivity was 52%. There were five cases that were identified during pre-surgical evaluation to depict cortical thickening. Considering solely these cases, 4/5 were correctly colocalized. The implication of this is that solely using cortical thickness as a measure in pre-surgical MRI examination, 5/31 cases would have the lesion identified. On the other hand, using our approach described in 3.1.1, 16/31 cases would have lesions selected. Thus there is an increased lesion detection rate using computational methods. However the lesional sensitivity was 33% and the lesional specificity was 85%. This indicates that while there was proper localization and low false detection rates, there was also an issue of limited coverage.

In Section 3.1.2, we examined identifying epileptogenic lesions using the gradient magnitude to monitor blurring of interface volumes with thresholding. The resulting subjectwise sensitivity was 68% such that 21/31 cases are correctly colocalized by this

method. The lesional sensitivity was 68% and lesional specificity 63%. While use of the interface improved coverage of the lesion, it was at the expense of additional structures being classified as false positives. 23/31 subjects were colocalized using both 3.1.1 and 3.1.2 methods with 9 of the cases being identified by one of the methods and the remaining 14 being identified in both methods. While both methods identified the majority of lesions, the fact that some cases were only picked up by a single method supports the notion of combining features detection to avoid inevitable false negatives.

A supplementary experiment in Section 3.1.3 considered the impact of utilizing volumetric T1-weighted images versus T2-weighted or FLAIR sequence acquisitions for examining the blurring of the gray-white interface. The respective subjectwise sensitivity for T2 and FLAIR was 74% and 81%. While the lesional sensitivity of these methods was on par with implementations with volumetric T1-weighted images, there outperformed in terms of lesional specificity.

Following preliminary morphometry experiments, classification techniques considering the structural volumes and texture attribute volumes were applied (Section 3.2). These experiments incorporated healthy controls and MRI-Positive subjects for training and performed subject based leave-one-out validation. In Section 3.2.6, we evaluated the performance of the various implementations relative to one another. While the 2-Step SVM implementation did not out-perform any other classifier method in terms of computed sensitivity or specificity values, using the ROC curves in combination with the modified J index resulted in the largest threshold selected. The implication of this is that the lesional clusters selected by the method are the largest (least susceptible to noise). The implementation using the Random Forest is noteworthy as it has the highest lesional sensitivity (81%) and the largest subjectwise sensitivity (94%). This implies that it offers the best coverage of lesions. However the method also had the lowest lesional specificity (67%) amongst all methods, implying that complete lesion coverage is at a cost of higher false positives. The implementation utilizing the 2-Step Naive Bayes classifier ultimately

produced the optimal results when considering the subjectwise specificity (100%), subjectwise sensitivity (94%) and lesional specificity (75%). It also had the shortest runtime (0.2 minutes). For this reason, the 2-Step Naive Bayes method was also applied to a set of four MRI-negative subjects in Section 3.2.7. The subjectwise sensitivity, lesional specificity, and lesional sensitivity were 50%, 25% and 94%. Thus, using classification techniques, lesions were selected in two of the four subjects that were not selected for in traditional MRI evaluation. Solely considering the results of the two correctly selected subjects, the associated lesional specificity, and lesional sensitivity were 60% and 87%. The low lesional specificity when considering all four cases is linked to differences in the sizes of lesions in the subjects that were missed. This issue is further discussed in future work. A final supplementary experiment comparing the implementation of the 2-Step Naive Bayes classifier method on volumetric T1-weighted, T2-weighted and FLAIR modalities was presented in Section 3.2.8, and the implementation on volumetric T1-weighted images was found to offer the best results in this case.

Comparing the results of classification methods to the previous preliminary experiments using thresholding, they had the opposite issue. The lesions had coverage but they were not localized. The classification techniques performed well in terms of selecting the correct structure but the issue was in over-selection. In adult cohorts for a dataset of 18 FCD cases (11 MRI-Positive and 7 MRI-Negative) and 14 healthy controls the resulting subjectwise specificity=100% and subjectwise sensitivity=83% using a 2-step Naive Bayes classifier [ACB$^+$03]. In the pediatric study we conducted for a dataset of 35 FCD cases (31 MRI-Positive and 4 MRI-Negative) and 13 healthy controls, the resulting subjectwise specificity=100% and subjectwise sensitivity=89% using the same method, indicating that the technique performs similarly on developing brains. Incorporating computational techniques such as this one can assist inexperienced radiologists in their detection, or provide initial hints for experienced radiologists on where to look for the lesion.

4.2 Future Work

In cases where the lesion is not restricted within a structure, the method faces difficulties. Manual lesion selection would likely improve results, particularly in regards to lesional sensitivity, however this is a time consuming task, as an expert would need to go through slice by slice to select the lesion. An alternate approach to this issue would be to incorporate some measure of spatial spread of selected lesional voxels. The number of healthy controls available also created an issue as it was too small to split into training and testing, and were therefore solely used for training. This likely contributes to the high subjectwise specificity seen across techniques. Future work can explore the inclusion of additional control cases.

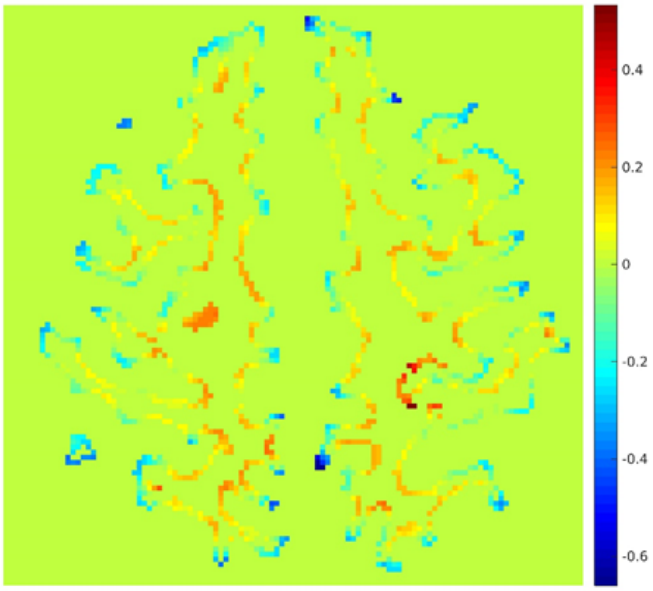
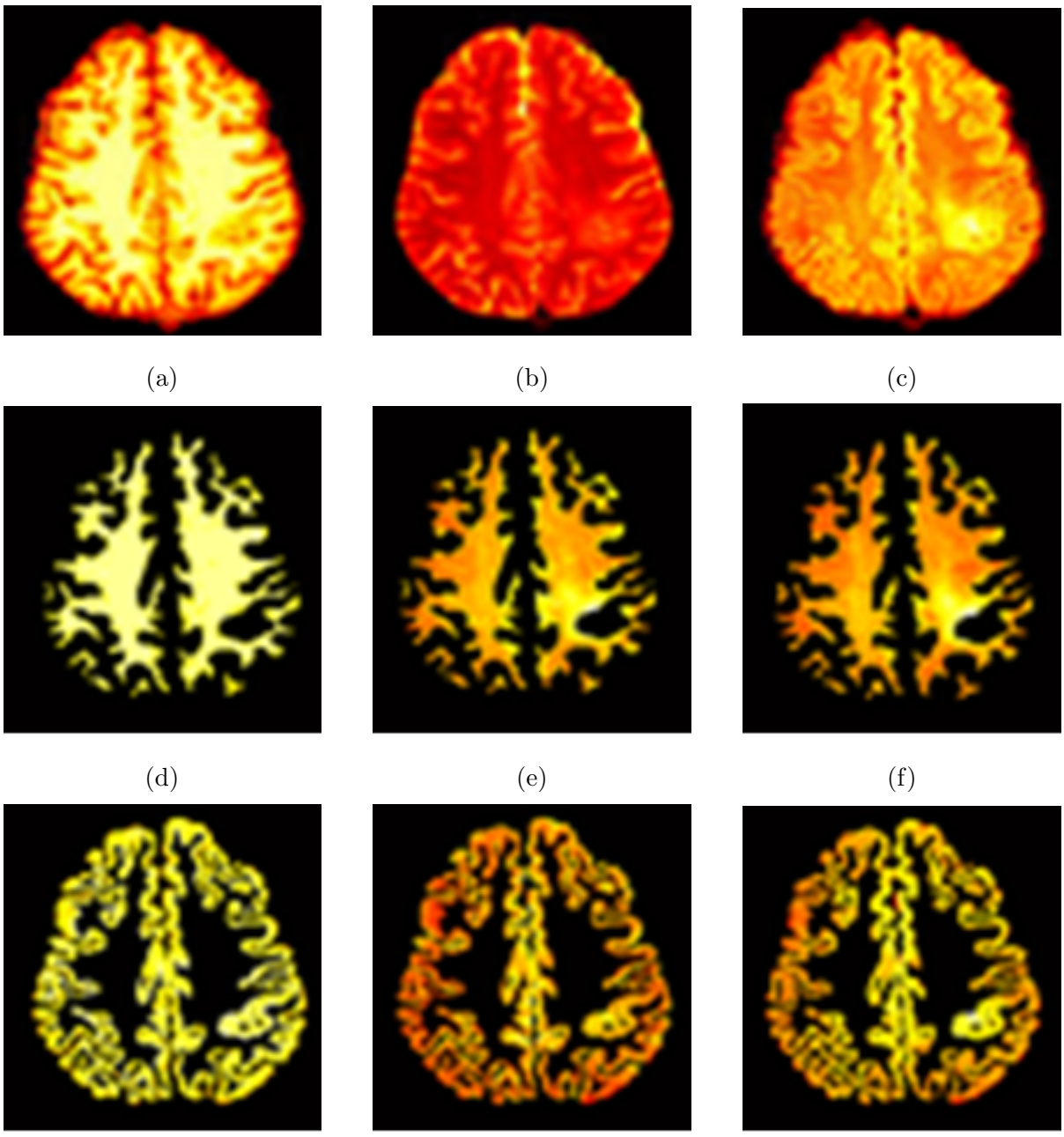


Figure 4.1: Axial slice of sulcal volume generated from sulcal surface. Gives an indication of linear distance and displacements, or depth of brain folds. Gyri are stored with negative values and sulci with positive values.

Two of the established MRI morphometry features were explored (cortical thickness and blurring of the gray/white matter interface). Although textural analysis has

been linked to detecting abnormal signals, this morphometric feature was not specifically developed into a volumetric model. Gray and white matter could be segmented and normalized individually to generate volumes for detection of irregular signals (Figure 4.2). Future work could also directly incorporate the remaining morphometry MRI feature, namely abnormal sulcal pattern, into the model. A possible method to examine sulcal behaviour could utilize the sulcal surfaces generated by FreeSurfer (Figure 4.1). Similar to the method for cortical thickness where thickness surfaces were used to make thickness volumes, in this case, the sulcal files would be used to produce sulcal fold volumes. [IJM+09] described this approach of using sulcal files from Freesurfer with the watershed algorithm to extract sulcal pits as a means of evaluating sulcal folds. Following this method would allow for easy integration into the classifier methodology we followed. Additionally, the method conducted for preliminary cortical thickness experiments and later incorporated into volumes for classification involved selecting vertices from confidence masks considering each slice along the axial axis. A modification to this method would be to also consider each slice along the coronal axis and then along the sagittal axis, before finally combining the results of all 3 directions. The same could be done for volumes produced for blurring of the interface.

Developing the experimental setup for classification followed methodology for sampling rates described by [ACB+03]. Given that we are dealing with an unbalanced 2 class problem, the argument could be made for other methods of synthetic sampling such as SMOTE (Synthetic Minority Over-sampling Technique) [CBHK02] or under-sampling to produce balance between the classes (lesional and non-lesional). When a small percent of patients are in the minority class, classifiers are more sensitive to detecting the majority class and less sensitive to the minority class. This causes the classifier output to be biased towards predicting the majority.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 4.2: Segmenting and normalizing gray and white matter for abnormal signal detection. (a) Original T1-weighted, (b) Original T2-weighted, (c) Original FLAIR, (d) T1-weighted white matter normalized, (e) T2-weighted white matter normalized, (f) FLAIR white matter normalized, (g) T1-weighted gray matter normalized, (h) T2-weighted gray matter normalized, (i) FLAIR gray matter normalized

The training is the most important step in the classifier's performance, thus it would be useful to explore modifications to this step. On that note, incorporating some method of feature selection is another possible modification. Some ideologies like decision trees naturally perform feature selection within the method. However, this is not the case for techniques like k-NN, which rely heavily on being provided useful information at the start. Selecting individual features that provide more information as opposed to all of them can possibly improve results. Other features to include in the future are Difference Entropy which was shown in studies to perform well in detecting lesions.

Previous work in computer aided lesion detection have predominantly explored T1-weighted images with some work in FLAIR, due to the thinner slices in T1 offering more detail. While the work in Section 3.1.3 indicated T2 and FLAIR performed better using thresholding, in Section 3.2.8 volumetric T1 outperformed in classification. Combining information from multiple modalities into a unified model for classification would be a possible consideration in the future.

The size of structures differed from cases to case and this affected performance evaluation. Lesional specificity is linked to the number of structures that the lesion falls within. In some cases the lesion is only contained in a single structure label (i.e. NL005), however for cases with many structures (i.e. Subject NL029 had 10 locations) failing to completely select every structure would negatively contribute to lesional sensitivity. Other methods have ranked selected regions on the basis of size and suspicion. Incorporating ranking into cluster output would offer a solution to dealing with the relatively lower lesional specificity value in the 2-Step Naive Bayes Classifier.

FCD is characterized by localized variations, which was one of the reasons that even in preliminary morphometry experiments with thresholding, there were satisfactory results. Current classification implementations [AWG+17] are leaning towards techniques involving localized feature examination. Modifying the implementation so that the classifier considers subsets of regions at a time ("searchlight method") is another variation on

the method that could be explored. The idea of exploring a modified version of this was considered using structural information, however there were issues in the balance of the structures. In some cases only one subject had a lesion in a particular structure, so using a leave-one-out approach meant that it would never be identified as lesional. However many cases had a lesion in the same structure, thus this region would be over-selected. For this reason it was not further explored in this study. On the other hand utilizing a moving region of interest would not necessitate a balance in the structure labels.

SIFT (Scale Invariant Feature Transform) [Low99], although not specifically used for FCD, has been promising for brain tumor detection [SH07]. The SIFT feature vector is calculated from the neighborhood around a keypoint at different scales. For this reason, localized structure is precisely recorded by SIFT. Due to the subtleties present in FCD, it would be an interesting topic to explore or incorporate into classification. However, manual tracing of lesions locations would likely be needed for the best results. This method utilizes a one-class method of classification, also performed by [EAHCL13]. We approached the problem as a two-class problem, however a one-class approach is another area that could be examined. In particular, doing so and including MRI-negative cases in the training data for classifying MRI-positive cases is a possibility in future work. With the exception of 3.2.7, all experiments solely used MRI-Positive subjects. Incorporating additional MRI-negative cases into training may contribute more in terms of recognizing subtle deviations in abnormal cases and improving results.

Bibliography

[A+08] Fritz Albregtsen et al. Statistical texture measures computed from gray level coocurrence matrices. *Image processing laboratory, department of informatics, University of Oslo*, 5, 2008.

[ACB+03] Samson B Antel, D Louis Collins, Neda Bernasconi, Frederick Andermann, Rajjan Shinghal, Robert E Kearney, Douglas L Arnold, and Andrea Bernasconi. Automated detection of focal cortical dysplasia lesions using computational models of their MRI characteristics and texture analysis. *Neuroimage*, 19(4):1748–1759, 2003.

[AWG+17] Sophie Adler, Konrad Wagstyl, Roxana Gunny, Lisa Ronan, David Carmichael, J Helen Cross, Paul C Fletcher, and Torsten Baldeweg. Novel surface features for automated detection of focal cortical dysplasias in paediatric epilepsy. *NeuroImage: Clinical*, 14:18–27, 2017.

[BAC+01] Andrea Bernasconi, Samson B Antel, Donald L Collins, Neda Bernasconi, Andre Olivier, François Dubeau, Gilbert B Pike, Frederick Andermann, and Douglas L Arnold. Texture analysis and morphological processing of magnetic resonance imaging assist detection of focal cortical dysplasia in extra-temporal partial epilepsy. *Annals of neurology*, 49(6):770–775, 2001.

[BBC+08] Pierre Besson, Neda Bernasconi, Olivier Colliot, Alan Evans, and Andrea Bernasconi. Surface-based texture and morphological analysis detects sub-

tle cortical dysplasia. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008*, pages 645–652, 2008.

[Bon80] A Bonis. Long term results of cortical excisions based on stereotactic investigations in severe, drug resistant epilepsies. In *Advances in Stereotactic and Functional Neurosurgery 4*, pages 55–66. Springer, 1980.

[Bre01] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[CBHK02] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[CJGM00] A Coeytaux, P Jallon, B Galobardes, and A Morabia. Incidence of status epilepticus in french-speaking switzerland (epistar). *Neurology*, 55(5):693–697, 2000.

[CNP+06] Richard FM Chin, Brian GR Neville, Catherine Peckham, Helen Bedford, Angela Wade, Rod C Scott, et al. Incidence, cause, and short-term outcome of convulsive status epilepticus in childhood: prospective population-based study. *The Lancet*, 368(9531):222–229, 2006.

[CTG+03] Nadia Colombo, Laura Tassi, Carlo Galli, Alberto Citterio, Giorgio Lo Russo, Giuseppe Scialfa, and Roberto Spreafico. Focal cortical dysplasias: MR imaging, histopathologic, and clinical correlations in surgically treated patients with epilepsy. *American Journal of Neuroradiology*, 24(4):724–733, 2003.

[DCDDS12] L De Cocker, F D'Arco, Ph Demaerel, and R Smithuis. Role of MRI in epilepsy. *The Radiology Assistant*, 1, 2012.

[DFS99] Anders M Dale, Bruce Fischl, and Martin I Sereno. Cortical surface-based analysis: I segmentation and surface reconstruction. *Neuroimage*, 9(2):179–194, 1999.

[DHS12] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. John Wiley & Sons, 2012.

[DS93] Anders M Dale and Martin I Sereno. Improved localizadon of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. *Journal of cognitive neuroscience*, 5(2):162–176, 1993.

[Dun97] John S Duncan. Imaging and epilepsy. *Brain: a journal of neurology*, 120(2):339–377, 1997.

[EAHCL13] Meriem El Azami, Alexander Hammers, Nicolas Costes, and Carole Lartizien. Computer aided diagnosis of intractable epilepsy with MRI imaging based on textural information. In *Pattern Recognition in Neuroimaging (PRNI), 2013 International Workshop on*, pages 90–93. IEEE, 2013.

[EAHJ+16] Meriem El Azami, Alexander Hammers, Julien Jung, Nicolas Costes, Romain Bouet, and Carole Lartizien. Detection of lesions underlying intractable epilepsy on T1-weighted MRI as an outlier detection problem. *PloS one*, 11(9):e0161498, 2016.

[EAP11] Haytham Elghazel, Alex Aussem, and Florence Perraud. Trading-off diversity and accuracy for optimal ensemble tree selection in random forests. In *Ensembles in Machine Learning Applications*, pages 169–179. Springer, 2011.

[EDA+15] Christoph Georg Eichkitz, John Davies, Johannes Amtmann, Marcellus Gregor Schreilechner, and Paul de Groot. Grey level co-occurrence matrix and its application to seismic data. *first break*, 33(3):71–77, 2015.

[Eke12] Suzanne Ekelund. ROC curves—what are they and how are they used? *Point of Care*, 11(1):16–21, 2012.

[FD00] Bruce Fischl and Anders M Dale. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*, 97(20):11050–11055, 2000.

[FLD01] Bruce Fischl, Arthur Liu, and Anders M Dale. Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE transactions on medical imaging*, 20(1):70–80, 2001.

[FSB+02] Bruce Fischl, David H Salat, Evelina Busa, Marilyn Albert, Megan Dieterich, Christian Haselgrove, Andre Van Der Kouwe, Ron Killiany, David Kennedy, Shuna Klaveness, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341–355, 2002.

[FSBD08] Niels K Focke, Mark R Symms, Jane L Burdett, and John S Duncan. Voxel-based analysis of whole brain flair at 3T detects focal cortical dysplasia. *Epilepsia*, 49(5):786–793, 2008.

[FSD99] Bruce Fischl, Martin I Sereno, and Anders M Dale. Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. *Neuroimage*, 9(2):195–207, 1999.

[FSvdK+04] Bruce Fischl, David H Salat, André JW van der Kouwe, Nikos Makris, Florent Ségonne, Brian T Quinn, and Anders M Dale. Sequence-independent

segmentation of magnetic resonance images. *Neuroimage*, 23:S69–S84, 2004.

[FVDKD+04] Bruce Fischl, André Van Der Kouwe, Christophe Destrieux, Eric Halgren, Florent Ségonne, David H Salat, Evelina Busa, Larry J Seidman, Jill Goldstein, David Kennedy, et al. Automatically parcellating the human cerebral cortex. *Cerebral cortex*, 14(1):11–22, 2004.

[GW02] Rafael C Gonzalez and Richard E Woods. Digital image processing using MATLAB, 2002.

[HB07] Mryka Hall-Beyer. The GLCM tutorial home page. *Current Version*, 2, 2007.

[HBS+16] Seok-Jun Hong, Boris C Bernhardt, Dewi S Schrader, Neda Bernasconi, and Andrea Bernasconi. Whole-brain MRI phenotyping in dysplasia-related frontal lobe epilepsy. *Neurology*, 86(7):643–650, 2016.

[HDO+98] Marti A. Hearst, Susan T Dumais, Edgar Osuna, John Platt, and Bernhard Scholkopf. Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4):18–28, 1998.

[HDVA+08] Chloe Hutton, Enrico De Vita, John Ashburner, Ralf Deichmann, and Robert Turner. Voxel-based cortical thickness measurements in MRI. *Neuroimage*, 40(4):1701–1710, 2008.

[HGF+05] Hans-Jürgen Huppertz, Christina Grimm, Susanne Fauser, Jan Kassubek, Irina Mader, Albrecht Hochmuth, Joachim Spreer, and Andreas Schulze-Bonhage. Enhanced visualization of blurred gray–white matter junctions in focal cortical dysplasia by voxel-based 3D MRI analysis. *Epilepsy research*, 67(1):35–50, 2005.

[HHKT16] Li-Yu Hu, Min-Wei Huang, Shih-Wen Ke, and Chih-Fong Tsai. The distance function effect on k-nearest neighbor classification for medical datasets. *SpringerPlus*, 5(1):1304, 2016.

[HHZ06] Torsten Hothorn, Kurt Hornik, and Achim Zeileis. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical statistics*, 15(3):651–674, 2006.

[HJS+06] Xiao Han, Jorge Jovicich, David Salat, Andre van der Kouwe, Brian Quinn, Silvester Czanner, Evelina Busa, Jenni Pacheco, Marilyn Albert, Ronald Killiany, et al. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. *Neuroimage*, 32(1):180–194, 2006.

[HKS+14] Seok-Jun Hong, Hosung Kim, Dewi Schrader, Neda Bernasconi, Boris C Bernhardt, and Andrea Bernasconi. Automated detection of cortical dysplasia type II in MRI-negative epilepsy. *Neurology*, 83(1):48–55, 2014.

[HS+73] Robert M Haralick, Karthikeyan Shanmugam, et al. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6):610–621, 1973.

[HTF09] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. Overview of supervised learning. In *The elements of statistical learning*, pages 9–41. Springer, 2009.

[HV03] David J Hand and Veronica Vinciotti. Choosing k for two-class nearest neighbour classifiers with unbalanced classes. *Pattern recognition letters*, 24(9):1555–1562, 2003.

[HW88] M Hajek and HG Wieser. Extratemporal, mainly frontal, epilepsies: surgical results. *Journal of Epilepsy*, 1(3):103–119, 1988.

[IJM+09] Kiho Im, Hang Joon Jo, Jean-François Mangin, Alan C Evans, Sun I Kim, and Jong-Min Lee. Spatial distribution of deep sulcal landmarks and hemispherical asymmetry on the cortical surface. *Cerebral cortex*, 20(3):602–611, 2009.

[Ize08] Alan Julian Izenman. *Modern multivariate statistical techniques*, volume 1. Springer, 2008.

[JCG+06] Jorge Jovicich, Silvester Czanner, Douglas Greve, Elizabeth Haley, Andre van der Kouwe, Randy Gollub, David Kennedy, Franz Schmitt, Gregory Brown, James MacFall, et al. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. *Neuroimage*, 30(2):436–443, 2006.

[JKS95] Ramesh Jain, Rangachar Kasturi, and Brian G Schunck. *Machine vision*, volume 5. McGraw-Hill New York, 1995.

[JMN13] Jana Jarecki, Björn Meder, and Jonathan D Nelson. The assumption of class-conditional independence in category learning. In *CogSci*, 2013.

[JNB+07] Lara E Jeha, Imad Najm, William Bingaman, Dudley Dinner, Peter Widdess-Walsh, and Hans Lüders. Surgical outcome and prognostic factors of frontal lobe epilepsy surgery. *Brain*, 130(2):574–584, 2007.

[KCB+03] T Kral, H Clusmann, I Blümcke, R Fimmers, B Ostertun, M Kurthen, and J Schramm. Outcome of epilepsy surgery in focal cortical dysplasia. *J Neurol Neurosurg Psychiatry*, 74(2):183–188, 2003.

[KGL16] Lohith G Kini, James C Gee, and Brian Litt. Computational analysis in epilepsy neuroimaging: A survey of features and methods. *NeuroImage: Clinical*, 11:515–529, 2016.

[KHD+07] Pavel Krsek, Milan Hajek, Monika Dezortova, Filip Jiru, Antonin Skoch, Petr Marusic, Josef Zamecnik, Martin Kyncl, Michal Tichy, and Vladimir Komarek. 1H MR spectroscopic imaging in patients with MRI-negative extratemporal epilepsy: correlation with ictal onset zone and histopathology. *European radiology*, 17(8):2126–2135, 2007.

[Lev15] Natalia Levshina. *How to do linguistics with R: Data exploration and statistical analysis*. John Benjamins Publishing Company, 2015.

[LLK+05] Sang Kun Lee, Seo Young Lee, Kwang-Ki Kim, Kkeun-Sik Hong, Dong-Soo Lee, and Chun-Kee Chung. Surgical outcome and prognostic factors of cryptogenic neocortical epilepsy. *Annals of neurology*, 58(4):525–532, 2005.

[Low99] David G Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee, 1999.

[LWN08] Christian Loyek, Friedrich G Woermann, and Tim W Nattkemper. Detection of focal cortical dysplasia lesions in MRI using textural features. *Bildverarbeitung für die Medizin 2008*, pages 432–436, 2008.

[LWS13] Falk Lüsebrink, Astrid Wollrab, and Oliver Speck. Cortical thickness determination of the human brain using high resolution 3T and 7T MRI data. *Neuroimage*, 70:122–131, 2013.

[MBF15] Pascal Martin, Benjamin Bender, and Niels K Focke. Post-processing of structural MRI for individualized diagnostics. *Quantitative imaging in medicine and surgery*, 5(2):188, 2015.

[MBR+07] Aileen McGonigal, Fabrice Bartolomei, Jean Régis, Maxime Guye, Martine Gavaret, Agnès Trébuchon-Da Fonseca, Henry Dufour, Dominique

Figarella-Branger, Nadine Girard, Jean-Claude Péragut, et al. Stereoelectroencephalography in presurgical assessment of MRI-negative epilepsy. *Brain*, 130(12):3169–3183, 2007.

[MGU02] Bruce McCune, James B Grace, and Dean L Urban. *Analysis of ecological communities*, volume 28. MjM software design Gleneden Beach, OR, 2002.

[Mit05] T Mitchell. Machine learning, generative and discriminative classifiers: Naive bayes and logistic regression (draft version), 2005.

[MSD81] Ryszard S Michalski, Robert E Stepp, and Edwin Diday. A recent advance in data analysis: Clustering objects into classes characterized by conjunctive concepts. 1981.

[Mur12] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[Mwi13] Kassim S Mwitondi. Data mining with rattle and R, 2013.

[OPB12] Thais Mayumi Oshiro, Pedro Santoro Perez, and José Augusto Baranauskas. How many trees in a random forest? In *MLDM*, pages 154–168. Springer, 2012.

[Per07] David Perlman. Cortical thickness: Practicalities and comparisons. *University of Wisconsin Statistics*, 692, 2007.

[PVL+14] Chintan Parmar, Emmanuel Rios Velazquez, Ralph Leijenaar, Mohammed Jermoumi, Sara Carvalho, Raymond H Mak, Sushmita Mitra, B Uma Shankar, Ron Kikinis, Benjamin Haibe-Kains, et al. Robust radiomics feature quantification using semiautomatic volumetric segmentation. *PloS one*, 9(7):e102107, 2014.

[Ras63] Theodore Rasmussen. Surgical therapy of frontal lobe epilepsy. *Epilepsia*, 4(1-4):181–198, 1963.

[RL01] Felix Rosenow and Hans Lüders. Presurgical evaluation of epilepsy. *Brain*, 124(9):1683–1700, 2001.

[Rom08] Bart M Haar Romeny. *Front-end vision and multi-scale image analysis: multi-scale computer vision theory and applications, written in mathematica*, volume 27. Springer Science & Business Media, 2008.

[RRF10] Martin Reuter, H Diana Rosas, and Bruce Fischl. Highly accurate inverse consistent registration: a robust approach. *Neuroimage*, 53(4):1181–1196, 2010.

[RSRF12] Martin Reuter, Nicholas J Schmansky, H Diana Rosas, and Bruce Fischl. Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage*, 61(4):1402–1418, 2012.

[SC08] Ingo Steinwart and Andreas Christmann. *Support vector machines*. Springer Science & Business Media, 2008.

[SDB+04] Florent Ségonne, Anders M Dale, Evelina Busa, Maureen Glessner, David Salat, Horst K Hahn, and Bruce Fischl. A hybrid approach to the skull stripping problem in MRI. *Neuroimage*, 22(3):1060–1075, 2004.

[SH07] Dong M. Hua J. Samuel, J. and E. M. Haacke. Brain tumor detection using scale invariant feature transform, 2007.

[SKO14] Katie Lee Salis, Soren Kliem, and K Daniel O'Leary. Conditional inference trees: a method for predicting intimate partner violence. *Journal of marital and family therapy*, 40(4):430–441, 2014.

[SPST+01] Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, and Robert C Williamson. Estimating the support of a high-dimensional distribution. *Neural computation*, 13(7):1443–1471, 2001.

[SS02] Bernhard Schölkopf and Alexander J Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.

[Suz12] Kenji Suzuki. Pixel-based machine learning in medical imaging. *Journal of Biomedical Imaging*, 2012:1, 2012.

[TJ+93] Mihran Tuceryan, Anil K Jain, et al. Texture analysis. *Handbook of pattern recognition and computer vision*, 2:235–276, 1993.

[WBF+01] HG Wieser, WT Blume, D Fish, E Goldensohn, A Hufnagel, D King, MR Sperling, H Lüders, and Timothy A Pedley. Proposal for a new classification of outcome with respect to epileptic seizures following epilepsy surgery. *Epilepsia*, 42(2):282–286, 2001.

[WCK+98] Elaine Wyllie, Youssef G Comair, Prakash Kotagal, Juan Bulacio, William Bingaman, and Paul Ruggieri. Seizure outcome after epilepsy surgery in children and adolescents. *Annals of neurology*, 44(5):740–748, 1998.

[You50] William J Youden. Index for rating diagnostic tests. *Cancer*, 3(1):32–35, 1950.

[ZRP+11] HB Zaman, P Robinson, M Petrou, P Olivier, TK Shih, SA Velastin, and I Nystrom. Visual informatics: Sustaining research and innovations. *LNCS, Springer. ISBN*, pages 978–3, 2011.

[ZS10] Heping Zhang and Burton Singer. *Recursive partitioning and applications*. Springer Science & Business Media, 2010.

[ZW09] Heping Zhang and Minghui Wang. Search for the smallest random forest. *Statistics and its Interface*, 2(3):381, 2009.