MODELLING THE VARIABILITY OF BRAIN STRUCTURES, ACTIVATION, AND CONNECTIVITY

A thesis
submitted to
the Nanyang Technological University

---

in fulfillment of the requirements
for the degree of
Doctor of Philosophy

Juan Zhou

October 2007

Division of Information Systems, School of Computer Engineering
Nanyang Technological University
Singapore

Acknowledgements

First and foremost, I would like to sincerely thank my supervisor, Associate Professor Jagath C. Rajapakse, for his support over the years. Besides his invaluable guidance and advice for my research, he provided so much freedom for me to explore and discover. He showed me different ways to approach a research problem and the need to be persistent in order to accomplish my goals. Without his encouragement and constant help, I could not have finished this thesis.

I would like to express my gratitude to all my friends and colleagues at BioInformatics Research Center of NTU with whom I have had the pleasure of working over the years. Special thanks go to Dr. Wang Yang, Dr. Zheng Xuebin, Dr. Tan Choong Leong, Dr. Duan Kaibo, Dr. Nguyen Ngoc Minh, Dr. Ho Sy Loi, Dr. Susanta Mukhopadhyay, Mr. Zheng Yun, Mr. Zheng Bo, Miss Tham Wei Ping Wendy, Miss Keren-Happuch E, and Mr. Yang Xiao for their great and honest discussions, comments and help all along the way. I also would like to send my appreciation to the executive officer, Ms. Norhana Binte Ahmad, and technicians, Mr. Tan Sing Yau and Mr. Kesavan Asaithambi, for their kind assistance.

Moreover, I would like to thank my parents and my beloved husband Bai Yan, for their constant support and encouragement throughout the period.

Lastly, many thanks to School of Computer Engineering, Nanyang Technological University for awarding the research scholarship and providing first-rate research facilities.

Table of Contents

Abstract

Magnetic resonance (MR) imaging techniques have revolutionalized *in vivo* investigation of the human brain through anatomical MR imaging (MRI) and functional MR imaging (fMRI). MRI is increasingly being used for quantification of brain tissues and structures while fMRI experiments are used to find brain regions and neural systems involved in a functional task. Due to the close relationship between structure and function of the brain, this thesis explores the variability of brain structures, activations, and connectivity across individual subjects in order to accurately analyze MRI/fMRI data.

Though numerous methods have been proposed for accurate identification of brain structures, none of the techniques have adequately addressed the variability of brain structures among individuals. We develop a novel fully automated approach based on fuzzy templates for segmentation of sub-cortical structures, taking into consideration their variability among individuals. Prior knowledge of structure intensity, spatial location, and relative spatial relationships are extracted automatically as fuzzy templates from a training dataset. The structures in a given brain image are then segmented using the fuzzy templates. The promise of the method is demonstrated in the segmentation of five sub-cortical brain structures.

The fact that the hemodynamic response function (HRF) varies across different regions and subjects has been seldomly addressed by earlier methods of detection of

brain activation from fMRI data. We propose a novel temporal-sliding-windows approach to extract simple and discriminating features from fMRI time-series to better handle hemodynamic variations. Fuzzy features thus derived are used to build fuzzy feature models by incremental learning. Furthermore, a novel weighted contextual fuzzy c-means clustering algorithm is proposed to detect brain activation by incorporating neighborhood information adaptively. The experiments on both synthetic and real fMRI data demonstrate that the proposed two approaches are capable of handling the variability of HRF across subjects and are less vulnerable to noise and artifacts.

We propose a method to model the variability of effective connectivity of the brain. Dynamic Bayesian networks (DBN) are introduced to model effective connectivity of the brain in order to overcome the drawbacks of earlier methods such as SEM, GCM, DCM, and BN. The DBN approach is completely data-driven and is able to explicitly model temporal relationships among brain regions to infer causal effects. We also propose a probabilistic framework associated with DBN to find an optimal connectivity structure accounting for variability among subjects. Experiments on both synthetic and real fMRI data illustrate that the approach is robust to the variability of HRF and inter-subject variability of brain connectivity.

In summary, this thesis develops novel methods for the analyses of both structural and functional MRI by detecting and modelling the variability of brain structures, activation, and connectivity. Future research directions include whole brain segmentation, detection of functional activation by incorporating anatomical priors, inference and validation of brain effective connectivity by multi-modality fusion, and applications to brain disease studies.

Author's Publications

Journal Papers

1. **Juan Zhou** and Jagath C. Rajapakse. Segmentation of subcortical brain structures using fuzzy templates. *NeuroImage*, 28(4), pp. 927-936, Dec. 2005.

2. **Juan Zhou** and Jagath C. Rajapakse. Modeling hemodynamic variability using fuzzy features for detecting brain activation from fMR time-series. *Neural Computing and Applications*, 16, pp. 541–549, 2007.

3. Jagath C. Rajapakse and **Juan Zhou**. Learning effective brain connectivity with Dynamic Bayesian Networks. *NeuroImage*, 37(3), pp. 749–760, Sep. 2007.

4. **Juan Zhou** and Jagath C. Rajapakse. Fuzzy technique to incorporate hemodynamic variability and contextual information for detecting brain activation from fMR images. *Neurocomputing*, 2008 (accepted).

5. Jagath C. Rajapakse, Yang Wang, Xuebin Zheng, and **Juan Zhou**. Probabilistic framework for detection of brain connectivity from functional MR Images. *IEEE Transactions on Medical Imaging*, 2008 (in press).

Other Journal Publications

1. **Juan Zhou**, Keren-Happuch E, Yih Yian Sitoh, Wei Ling Lee, Jih Jing Chin, Mei Sian Chong, and Jagath C. Rajapakse. Differences in neural correlates of visual spatial and visual object working memory: an fMRI study. *NeuroImage*, 31(Sup. 1), S59, June 2006 (abstract).

2. Keren-Happuch E, **Juan Zhou**, Yih Yian Sitoh, Wei Ling Lee, Jih Jing Chin, Mei Sian Chong, and Jagath C. Rajapakse. Attentional network involving dorsal ACC: digit-based counting Stroop fMR study. *NeuroImage*, 31(Sup.1), S102, June 2006 (abstract).

Book Chapters

1. **Juan Zhou** and Jagath C. Rajapakse. Extraction of fuzzy features for detecting brain activation from functional MR time-series. I. King et al., (Eds.), *Neural Information Processing, Lecture Notes in Computer Science, LNCS* 4234, pp. 983-992, Springer-Verlag Berlin Heidelberg, 2006.

2. Neelam Sinha, Manojkumar Saranathan, A. G. Ramakrishnan, **Juan Zhou**, and Jagath C. Rajapakse. Ultra-fast fMR imaging with high-fidelity activation map. I. King et al. (Eds.) *Neural Information Processing, Lecture Notes in Computer Science, LNCS* 4233, pp. 361-368, Springer-Verlag Berlin Heidelberg, 2006.

Conference Papers

1. **Juan Zhou** and Jagath C. Rajapakse. Decision-level fusion of structural and functional MR images. *Proceedings of the 12th International Conference on Neural Information Processing*, Oct 30 - Nov 2, 2005.

List of Tables

List of Figures

Abbreviations and Symbols

Abbreviations

3D	three-dimensional
4D	four-dimensional
AC	anterior commissure point
ACC	anterior cingulate cortex
AFG	anterior frontal gyrus
BA	Brodmann areas
BIC	Bayesian information criterion
BN	Bayesian networks
BOLD	blood-oxygenation-level-dependent
cFCM	contextual fuzzy c-means clustering
CPD	conditional probability density
CRF	conditional random field
CSF	cerebrospinal fluid
DAG	directed acyclic graph
DBN	dynamic Bayesian networks
DCM	dynamic causal modeling
DTI	diffusion tensor imaging
EC	extrastriate cortex
EEG	electroencephalography
EPI	echo planar imaging
FDR	false discovery rate
FFM	fuzzy feature modelling
FFX	fixed-effect analysis
FWHM	full width at half maximum
FLASH	fast low angle shot
fMRI	functional magnetic resonance imaging
FOV	field of view
FWER	familywise error rate
GCM	Granger causality mapping
GLM	general linear model
GRF	Gaussian random field

GM	gray matter
HRF	hemodynamic response function
ICA	independent component analysis
ICBM	international consortium for brain mapping
IFG	inferior frontal gyrus
IPL	inferior parietal lobe
LAS	radiological convention in brain orientation
MAP	maximum a posteriori
MCMC	Markov chain Monte Carlo
MFG	middle frontal gyrus
MH	Metropolis-Hastings
ML	maximum likelihood
MNI	Montreal Neurological Institute
MRF	Markov random field
MRI	magnetic resonance imaging
MTC	middle temporal cortex
PC	posterior commissure point
PCA	principal component analysis
PET	positron emission tomography
PMA	primary motor area
QHC	Quasi-Hemodynamic curve
RF	radio frequency
RFT	random field theory
RAS	neurological convention of brain orientation
SEM	structural equation modelling
SMA	supplementary motor area
SNR	signal to noise ratio
SPECT	single photon emission computed tomography
SPL	superior parietal lobe
SPM	statistical parametric mapping
TE	echo time of MR system
TR	repetition time of MR system
TSW	temporal-sliding-windows
VAR	vector autoregressive
VIFG	ventral inferior frontal gyrus
wcFCM	weighted contextual fuzzy c-means clustering
WM	white matter

Important Symbols

c_{ij}	VAR coefficient between region i and j
D	a dataset
f^j	structural MR image of subject j
e	the Gaussian innovation
G_j^l	a set of image voxels in structure l of subject j
I_1	percent volume difference of structure
I_2	percent volume overlap of structure
I_3	mean distances measure of structure
M	image transformation
N	the number of subjects in a dataset
$p(a\|b)$	the conditional probabilities of a given b
R	a set of brain regions
r	the standard brain structural template
S	a network structure
s^j	expert segmentation of structural MR image of subject j
T	the length of fMR time-series
$x_i(t)$	a random variable representing ith activated brain region at time t
X	design matrix in GLM or condition in a fMRI experiment
Y	fMRI time-series
Z	fuzzy membership map of activation
θ	the parameters of CPDs
Γ	a set of brain sub-cortical structures
Ω	a three-dimensional spatial domain of image voxels

Chapter 1

Introduction

1.1 Motivations

Over the past ten years, rapid improvements in imaging technology and methodology have enabled scientists to study human brain anatomy and function in detail. Magnetic resonance (MR) is one of the most important imaging techniques, which generate both structural MR images (MRI) and functional MR images (fMRI). The structural MR imaging technique is able to provide detailed knowledge of normal and diseased anatomy for medical research and has become a critical component in clinical diagnosis and treatment planning. Functional MR images allow an *in vivo* measure of brain activity, and are increasingly utilized to explore brain networks and neural interactions underlying brain functions [1]. Though a wide variety of methods have been developed to analyze these two types of images, extracting the most accurate structural and functional information of the human brain is still challenging as previous studies have not dealt with the inter-subject variability involved in data. Therefore, this thesis focuses on handling the variability of brain structures, activation, and connectivity across individual subjects in order to accurately analyze MRI/fMRI data. In order to accomplish this, novel methodologies to detect brain structures, activation, and connectivity are proposed. The following three problems

are mainly addressed: (1) segmentation of subcortical structures from structural MRI; (2) detection of functional activation from fMRI; and (3) inference of effective brain connectivity from fMRI.

In a functional imaging study, a series of fMR images are acquired together with a separate high-resolution structural image to help localizing functional activations of the human brain. However, identifying the location and activation of brain functions by fusing structural and functional images is both conceptually and technically more difficult and sensitive than has been actually assumed [2]. Because of the anatomical and functional variability among subjects, normalization to a standard brain template is usually adopted to compare signal locations among individual subjects and studies. Scientists usually overlay the activation pattern derived from fMRI over segmented structural MRI and report Talairach [3] or MNI (Montreal Neurological Institute) coordinates for regions of interest. Therefore, segmentation of structural MRI is an essential step to localize functional activations in human brain. Moreover, inference of effective brain connectivity from fMRI is a region-based approach, which makes use of cluster-wise average time-series in every brain region. The accurate localization of brain activation is a prerequisite for brain connectivity study. Therefore, detecting structures, activation, and connectivity are closely related in studying brain functions through imaging. It is of paramount importance to handle the variability of these three aspects in order to accurately quantify brain structures, activations, and connectivity.

Segmentation of MR images of the human brain can be divided into two types: tissue segmentation and structure segmentation (subcortical and cortex). In this thesis, the focus is on segmentation of subcortical structures because quantitative analyses of neuroanatomical or neurofunctional data often require explicit identification of boundaries of subcortical structures. The fundamental question is how to accurately segment different brain subcortical structures automatically from MRI with computer technology. The first issue is the inhomogeneous intensity within one tissue class or

structure and the overlapping of intensity characteristics among different structures, which reduces the effectiveness of intensity-based methods. Although the spatial information can be incorporated to solve the inherent ambiguities of class intensity distributions, due to the complexity and variability of the location, shape, size, and orientation of anatomical structures among individuals, accurate identification of subcortical structures poses many challenges. Other issues such as low contrast, noise perturbation, partial volume effect, and inhomogeneities of magnetic field in MR scanner should also be addressed [4]. A wide variety of methods have been proposed in the literature for segmentation of subcortical structures [5], including deformable models, atlas-based methods, and hybrid approaches. But most segmentation methods only generate a coarse location of the desired objects and nowadays clinicians still rely on manual segmentation of brain subcortical structures from MR images. There is a need to investigate novel techniques that can automatically detect subcortical structures more accurately.

Apart from studying brain anatomy using structural MRI, scientists are more and more interested to use functional MRI to study brain functions as the fMRI renders brain activity both spatially and temporally. The detection of fMRI response, however, is a non-trivial task because of the subtle signal changes due to stimuli and their unknown exact forms. Moreover, the variability of hemodynamic response function (HRF) among individuals and brain regions further confound the signal. This motivates the development of intelligent approaches to discover various response patterns and associated activated brain regions [6].

Generally, techniques for detection of functional activation from fMRI fall into two categories: hypothesis-driven and data-driven. Hypothesis-driven techniques are based on explicit prior knowledge about stimuli or activation patterns and currently the most widely used method is statistical parametric mapping (SPM) [7] derived from the well-known general linear model (GLM) [8]. However, these techniques are not

appropriate for studies in which the responses are unknown or differential HRFs are expected [9]. The validity of statistical approaches depends on the suitability of underlying assumptions of signals and noise to real data. Data-driven methods, in contrast, do not assume prior models but they have difficulty in results interpretation by physiological terms. Emerging studies involve more sophisticated cognitive tasks and the analyses of such experiments necessitate more intelligent techniques that do not require prior information or assumptions of the data, and at the same time, enable explicit interpretation of underlying physiology.

The brain regions involved in a certain cognitive function represent functional specificity which, nevertheless, does not reflect the underlying communications among brain regions. Modern neuroscience focuses on functional integration studies, i.e., how different parts of brain regions interact in order to perform a particular task. This provides a holistic view of brain functions, represented by brain connectivity, which has become one of the hot topics in brain mapping research. Effective connectivity defined as "the influence that one neural system exerts over another either directly or indirectly"by Friston [10] is an important measure of such brain connectivity. Structural equation modelling (SEM) was first proposed to infer brain connectivity [11], which only describes the neural system in second-order statistical sense. Dynamic causal modelling (DCM) was recently developed to deal with effective connectivity at neuronal level [12]. But both DCM and SEM are confirmatory techniques in the sense that they both require prior models or hypotheses to begin with, which are often under anatomical constraints and complicated by the fact that many of them have been obtained in the studies of monkeys. It is not always certain which areas are to be included in the study, especially if the brain regions are involved in functions unique to human, such as language and cognition [13]. Furthermore, the incomplete or incorrect specification of the prior models may lead to faulty inference of brain functions and connectivity.

Granger causality mapping (GCM) is an exploratory approach to investigate effective connectivity among activated brain areas using a vector autoregressive (VAR) model [14]. However, the dynamics of signal-fluctuations in GCM are modelled by random error process and the input stimuli are not explicitly incorporated as causes of fluctuations. Recently, Bayesian networks (BN) was proposed to explore the brain networks in a completely exploratory manner [15]. However, BN does not provide a direct mechanism to represent temporal dependencies among multiple processes at brain regions and therefore cannot decisively infer causal relationship among brain regions. Moreover, inter-subject variability in effective connectivity has seldom been addressed. Earlier methods either treat the group data as a single subject or find the connectivity for each subject separately. In summary, a completely data-driven approach with explicit modelling of temporal precedence is required to infer causal relationships among brain regions and a more intelligent technique should be developed to find an optimal connectivity structure for group studies considering the commonalities and differences among subjects.

1.2 Major Contributions

The major achievements of this thesis reside in the detection of brain structures, activations, and connectivity from structural and functional MR images, taking into account their variability among individuals. Specific contributions of the thesis are highlighted below:

- We proposed a novel method to automatically segment brain subcortical structures from magnetic resonance images. In order to address structural variability among individuals, a set of fuzzy templates of the structures based on features such as intensity, spatial location, and relative spatial relationships are created

from a set of training images automatically. Segmentation is then performed on test images using registered fuzzy templates and tissues maps. The method does not require specific expert definition of each structure, instead, prior knowledge is automatically extracted and represented from the training images. No manual initializations or interactions are required during segmentation process. The robustness of the method was demonstrated by segmenting five structures: thalamus, putamen, caudate, hippocampus, and amygdala. The performance is comparable with previous techniques. The effect of image registration was also studied.

- A novel fuzzy feature modelling (FFM) approach was developed to detect functional activation from fMRI. The method consists of (1) extracting features from fMR time-series automatically by a sequence of temporal-sliding-windows (TSW); (2) fuzzy feature modelling by incremental learning; and (3) thresholding on fuzzy activation map to obtain significant activation. The extracted features are less vulnerable to noise and lead to a lower dimensional space. Fuzzy modelling is able to handle complex, nonlinear, and high uncertain fMRI data, with the limited available information. Incremental learning adapts to the inter-subject variability of hemodynamic response function (HRF). The experiments on detection of functional activation from both synthetic and real fMRI data illustrate that the proposed approach is less vulnerable to correlated noise and is able to adapt to different hemodynamic response functions across subjects.

- In order to incorporate contextual constraints embedded in fMRI data adaptively, a new version of unsupervised weighted contextual fuzzy c-means clustering (wcFCM) was developed. The cost function of contextual FCM (cFCM) that is regularized by spatial context, is modified to selectively capture information from neighboring voxels based on its similarity to the reference voxel. Weighted contextual FCM is applied on the extracted feature space of fMRI to

identify activated voxels. Each subject is handled separately assuming no fixed structure of signal and noise to account for inter-subject variability of HRF. The experiments on detection of functional activation from both synthetic and real fMRI data show that the wcFCM is superior than SPM for handling the variability of noise and HRF.

- Dynamic Bayesian networks (DBN) was introduced to model effective connectivity of the brain in order to overcome the drawbacks of earlier methods including SEM, GCM, DCM, and BN. The DBN approach is completely data-driven (exploratory) in the sense that it does not require an *a priori* model as earlier methods. It is able to model temporal precedence among brain regions, meaning that causal relationships are represented explicitly. The conditional probabilities in the resulting networks represent interactions among brain regions in complete statistical sense. Moreover, a probabilistic framework associated with DBN was developed to find an optimal connectivity structure from group studies accounting for the variability across subjects. The method was tested and compared with earlier methods like GCM by experiments on synthetic data; the effects of sampling, noise, and the variability of HRF were evaluated. FMRI data collected in silent word reading and counting Stroop task were also used to find effective brain connectivity and the results were consistent with the literature and comparable to the results produced by BN method.

1.3 Organization of the Thesis

This thesis is organized in order to describe the project from basic background to specific topics in detail. It consists of six chapters.

Chapter 1 begins with a brief introduction on the background and motivations of this

thesis, followed by an outline of major contributions.

Chapter 2 gives the necessary knowledge of the topics discussed in this thesis including areas such as neurobiology, neuroimaging, and computational techniques. The basic principle of structural and functional MR images is first described. The three aspects of the human brain, i.e, anatomy, function, and connectivity, are then presented. The issues and earlier methods of three problems addressed in this thesis, that are, segmentation of subcortical structures from MRI, detection of functional activation from fMRI, and inference of effective connectivity from fMRI, are reviewed separately.

Chapter 3 presents the proposed approach to segment subcortical structures from MRI using fuzzy templates, emphasizing on how the technique accounts for inter-subject variability of structures. Segmentation of five subcortical structures from real MRI data are illustrated. The performance comparison with previous approaches is presented using manual segmentation as reference.

Chapter 4 describes two fuzzy techniques for detection of functional activation from functional MRI: fuzzy feature modelling (FFM) and weighted contextual fuzzy c-means clustering (wcFCM). The performance of these two methods are compared to SPM by experiments on both synthetic and real fMRI data. The results show that both approaches are able to adapt to the variability of HRF; in addition, wcFCM is able to capture temporal and spatial neighborhood information adaptively.

Chapter 5 introduces dynamic Bayesian networks (DBN) to infer effective brain connectivity from fMRI data. Experiments on synthetic and real fMRI data show that DBN is capable of modelling temporal relationships among brain regions to infer causal effects. The robustness of DBN is demonstrated by evaluations of the effects of sampling, noise, and HRF variability as well as comparisons with earlier approaches like GCM and BN. The extension of DBN for modelling inter-subject variability in effective connectivity is also described.

Finally, chapter 6 concludes this thesis by giving possible directions to extend the research done.

Chapter 2

Review of Previous Work

This chapter starts with the background of structural and functional magnetic resonance imaging and then of brain structures, function, and connectivity. Segmentation of brain subcortical structures is a critical task for medical diagnosis, for which earlier approaches like deformable models and atlas-based models are discussed. Secondly, previous work on the detection of brain activation from fMRI is reviewed with a focus on statistical parametric mapping (SPM) and data-driven methods. Lastly, three techniques for inference of effective brain connectivity namely, structural equation modelling (SEM), dynamic causal modelling (DCM), and Granger causality mapping (GCM), are presented.

2.1 Background

In this section, we first introduce the principles and related applications of structural and functional MRI. The anatomical and functional specificities of the human brain are described thereafter. A discussion on brain connectivity in terms of functional and effective connectivity is then followed.

2.1.1 Structural MRI

Nuclear magnetic resonance (NMR), a property of atoms first observed by Bloch and Purcell in 1946, has proven to be an informative technique in many fields. NMR is based on the vibration of atoms in a strong magnetic field. A nucleus of an atom has spin states of different energy levels, which is excited to make a transition from a low energy state to a high energy state or the other way around, if the atom absorbs or emits a photon whose energy matches the difference between the states.

Magnetic resonance imaging (MRI), an extension of NMR, enabled scientists to study human anatomy by generating high-contrast images based on different relaxation properties of proton NMR signal in various soft tissues and organs. MRI focuses on hydrogen atoms because of their high concentration and large magnetic moment. Like other tissues in the body, over 70 percent of the brain is composed of water. Based on the fact that different parts of the brain have slightly different amounts of water, MRI is capable of generating a contrast between the cortical surface and the underlying white matter. Thus, detailed brain anatomy including structures and tissues are examined through MRI. A typical MR system consists of the following components [16]:

- A large magnet to generate the magnetic field;
- Shim coils to make the magnetic field as homogeneous as possible;
- A radio frequency (RF) coil to transmit a radio signal into the body part being imaged;
- A receiver coil to detect the returning radio signal;
- Gradient coils to provide spatial localization of the signal;
- A computer to reconstruct the radio signal into the final image.

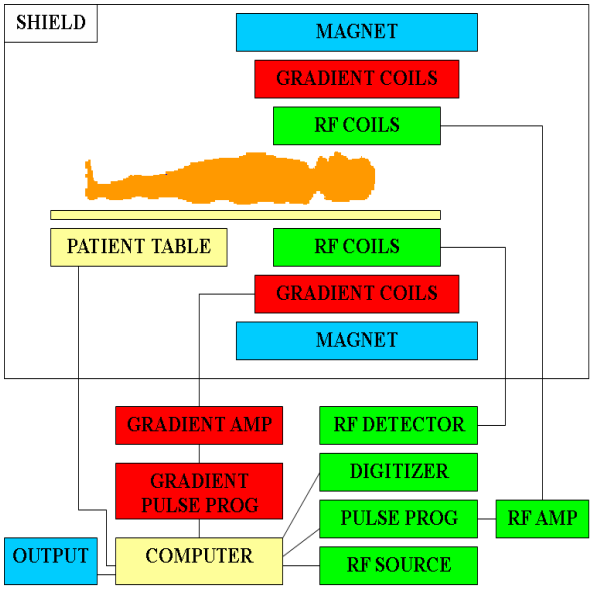


Figure 2.1: Hardware overview of a typical MR system [18]

A schematic representation of a typical MR system with its major components and interconnections are shown in Fig. 2.1. The basic physics behind a MR system is that protons in a magnetic field are capable of receiving and transmitting electromagnetic energy. And the strength of the transmitted energy is proportional to the number of protons in the tissue [17]. Thus, when a radio frequency (RF) pulse from the RF coil perturbs the steady-state proton magnetization, a small radio signal is emitted from protons, which is received by the receiver coil. Spatial encoding is done by gradient coils and final map of MR signal strength is reconstructed as an image on the computer.

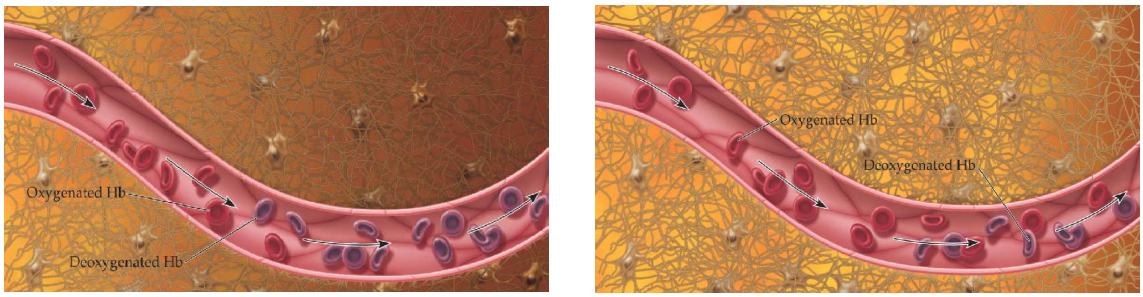
The signal intensity of MR image is determined by four basic parameters: proton density, T1 relaxation time, T2 relaxation time, and flow. Proton density is the concentration of protons of the tissue in the form of water and macromolecules, which is slightly different in different parts of the brain. The T1 and T2 relaxation times define the way that the protons revert back to their resting states after the initial RF pulse. The most common effect of flow is the loss of signal from rapidly flowing arterial blood. By changing the pulse sequence and imaging parameters, T1-weighted

(longitudinal relaxation time), T2-weighted (transverse relaxation time), and proton density-weighted MR images are produced. The pulse sequence sets the specific number, strength, and timing of RF and gradient pulses. Two most important parameters are repetition time (TR), the time between consecutive 90° RF pulses, and echo time (TE), the time between the initial 90° RF and the echo. Spatial resolution is determined by parameters like the matrix size, the field of view (FOV), and slice thickness. TR, the matrix size, and slice thickness influence scan time.

2.1.2 Functional MRI

Functional magnetic resonance imaging (fMRI), a variation of MRI, is a new and exciting technology used to measure brain activity *in vivo*. The report of the first fMRI was given by Ogawa et al. [19, 20]. Functional MRI is a noninvasive technique and provides unsurpassed spatial resolution than other functional neuroimaging modalities such as electroencephalography (EEG) and positron emission tomography (PET). FMRI is also less physically demanding on subjects than single photon emission computed tomography (SPECT) or PET which require injection of a small amount of radioactive substance. The following paragraphs will describe the principles of fMRI signal and common procedures in fMRI experiments.

Functional MRI is based on the assumptions that increased neural activity in one region leads to an increase in blood oxygenation and highly oxygenated neuronal populations produce a larger magnetic resonance signal, which is referred to as blood-oxygenation-level-dependent (BOLD) signal. The neurons in the brain consume oxygen when they act and hemoglobin molecules in the blood continuously provide oxygen to the neurons. When the neuronal activity increases, so does the demand of oxygen and an increase of flow of blood is seen in regional supply to the activated neuronal population. Fig. 2.2 illustrates that an excess of oxygen is supplied to the active neu-



(a) (b)

Figure 2.2: BOLD signal of fMRI [24]: (a) the baseline state; and (b) the active state, where more oxygenated blood (in red color) is observed.

rons, compared to (a) the baseline state, leading to an increase in the concentration of oxygenated blood in the capillaries surrounding the brain area in (b) the active state. This difference can be captured by a MR scanner due to different magnetic properties of oxygenated and deoxygenated blood.

BOLD signal is an indirect indicator of brain activity. Logothetis et al. [21] has done simultaneous measurements of electrical activity and blood oxygenation to show that electrical neuronal activity and BOLD signal are actually intimately coupled. It is impossible to measure the activity of single neuron by BOLD signal as a large population is required to evoke a measurable response. The typical spatial resolution down to 1mm can be achieved [22], yet in practice, a few millimeters resolution is used in fMRI experiments. In contrast to structural MRI, fMRI uses T2*-weighted imaging in a conventional MR machine. The gradient echo planar imaging (EPI) sequence is often used to generate functional MR images. EPI, devised by Mansfield [23], is the fastest MRI sequence available nowadays to capture an image slice in less than 100ms and an entire brain volume in just a few seconds. This provides fMRI with balancing spatial and temporal resolution to study brain activity.

With the advent of fMRI as a tool for mapping brain activity, the field of neuroscience has experienced a boost due to the ability to study human brain function in a non-intrusive manner. Numerous fMRI experiments have been performed to investigate

higher cognitive functions such as language and memory as well as the basic sensory, visual, and motor brain functions. The goal of a typical fMRI study is to functionally associate one or more brain regions with the task performed by a subject and to infer the possible underlying interaction mechanisms among brain regions. A schematic diagram of the general procedure of a fMRI study is given in Fig. 2.3. The study begins by designing an appropriate task to investigate a specific hypothesis on brain functions and then undergos a series of behavioral and imaging experiments. By analyzing the data collected from both experiments, relevant brain functional activation is localized and the corresponding structure of brain connectivity is inferred. Psychological interpretation of the results is required to ensure the validity of the proposed hypothesis.

The hypothesis on brain functions and the design of experimental tasks are the key ingredients in a fMRI study. Currently, there are two most commonly used experimental design techniques: block design (box-car) and event-related design. In a block design experiment, stimuli are usually presented at a fixed pace for a constant period, alternating with rest periods of fixed-length, in which no stimulus is presented. The duration of these blocks needs to be long enough to accommodate the hemodynamic response.

In contrast, event-related design does not necessarily have a fixed inter-stimulus length. Stimuli are presented based on a pre-designed protocol at a varying rate and are also subjected to the subject's performance in terms of response time and accuracy. Each trial is intermixed randomly with other types of trials to achieve statistical independence. While the block design has a strong detection power, the event-related design has the ability of estimating the shape of hemodynamic response function (HRF). Due to different HRFs corresponding to various physiological tasks [25], the objective of an event-related design is the eventual separation of responses of activation for different task. The total length of paradigm for the above two types

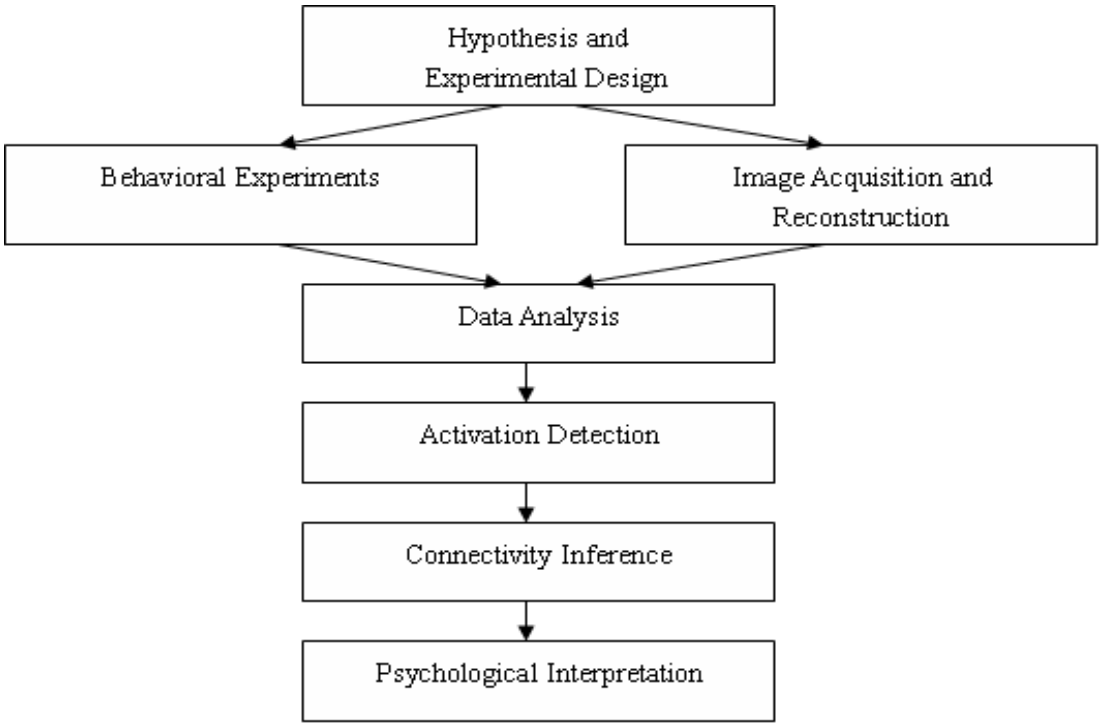


Figure 2.3: Illustration of a typical fMRI experimentation and analysis.

of design should be long enough to gain valid statistical power for the detection of activation. Nevertheless, a compromise has to be made between the comfortability of the subject and the number of scans needed to achieve sufficient signal to noise ratio. Another issue of fMRI is that the shape of HRF varies across subjects, regions, and tasks.

2.1.3 Brain Structure and Function

The human brain is an incredibly complex organ, probably most complex of all known objects in the universe, which we have only just begun to understand. The human brain can be divided into three major regions: forebrain, midbrain, and hindbrain. Hindbrain is the first evolved primitive brain, which still presents in humans today. The rest of the brain is connected to the spine via the hindbrain which is responsible for many life-sustaining functions. The midbrain allows humans to orient themselves

Major Divisions	Subdivisions	Structures
Forebrain	Telencephalon	Neocortex; basal ganglia; amygdala; hippocampus; lateral ventricles
	Diencephalon	Thalamus; hypothalamus; epithalamus; third ventricle
Midbrain	Mesencephalon	Tectum; tegmentum; cerebral aqueduct
Hindbrain	Metencephalon	Cerebellum; pons; fourth ventricle
	Myelencephalon	Medulla oblongata; fourth ventricle

Table 2.1: Divisions of the brain and its major subcomponents.

according to sensory stimuli. The forebrain includes the components that create basic and complex human behaviors [26]. Each region is further broken down into small structures with different functions. Table 2.1 presents the subcomponents of these major divisions of the human brain [27].

The brain matter can be partitioned into three types of tissues: gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). GM looks relatively dark in anatomical brain specimens due to densely packed cell bodies of neurons. GM is found in both cerebral cortex and subcortical regions where high concentration of neurons exists. Axons are projected out of neurons, which connect to other neurons and permit communications among neurons even at distant locations. WM appear white in both postmortem brain and T1-weighted MR images as it contains axons covered with a fatty insulation, myelin. CSF is a fluid containing nutrients and byproducts of brain activity and the brain is bathed inside and out by CSF. The regions that contain CSF inside the brain are named ventricles.

The brain is also divided at the middle lengthwise through corpus callosum into two hemispheres. Each hemisphere of the cerebral cortex is demarcated into four lobes by various sulci (groove) and gyri (bump). The folding of the cerebral cortex produced by gyri and sulci increases the area of the cerebral cortex that fits in the skull. No two brains have exactly the same shape of cerebral cortex [27]. Each of the four lobes is characterized by its location and specific functions. *Frontal lobe* is located

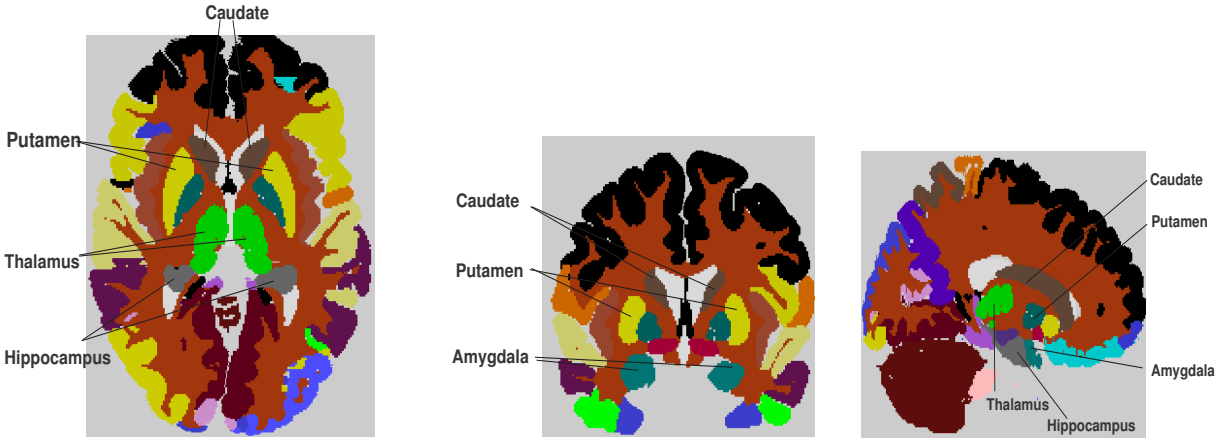
at the anterior end of each cerebral hemisphere and in front of the central sulcus, which is involved in reasoning, planning, part of speech and movement, emotions, and problem-solving. *Parietal lobe* is located at the top of each hemisphere and behind the central sulcus, which is in charge of perception of stimuli related to touch, pressure, temperature, and pain. *Temporal lobe* is located at the bottom of each hemisphere and below the lateral fissure, which participates in perception and recognition of auditory stimuli as well as memory. *Occipital lobe* is located at the back of the brain and behind the parietal lobe and temporal lobe, which is largely involved in many aspects of vision.

Besides cerebral cortex mentioned above, subcortical structures of the human brain are of essential functions. Among them, five subcortical structures interested in this thesis: thalamus, putamen, caudate, amygdala, and hippocampus, are described as below. The spatial locations of these structures are illustrated in Fig. 2.4 in three different views. It is worth emphasizing that inter-subject anatomical variability of these structures exist, in addition to developmental changes.

- *Thalamus* is a structure deep within the brain stem that receives sensory information from the nervous system and relays the information selectively to the cerebral cortex and other parts of the brain. Thalamus acts as the director of information related to bodily functions such as seeing, sleeping, hearing, waking, tasting, and touching;

- *Putamen* is a nucleus of the basal ganglia lateral to the globus pallidus, which, together with the caudate nucleus form the striatum. It is known to play a role in reinforcement learning and work with caudate to influence motor activity;

- *Caudate* is a structure within the basal ganglia and is responsible for regulating and organizing information being sent to the frontal lobes from other areas of the brain. Besides primarily involved in control of voluntary movement, it is



(a) Axial view (b) Coronal view (c) Sagittal view

Figure 2.4: Subcortical structures of the brain.

also a crucial part of the learning and memory system of the brain.

- *Amygdala* is an almond-shaped structure located deep within the medial temporal lobes of the brain. It is part of the limbic system, performing a primary role in the processing and memory of emotional reactions.

- *Hippocampus* is part of the limbic system deep in the temporal lobe and is strongly involved in learning, memory formation, and spatial navigation.

Since we are dealing with MR image of the brain, it is necessary to be familiar with the conventions and terminologies used to describe the orientation of the brain. The human brain is viewed as three-dimensional biological structure, whose orientation and direction are defined relative to the subject, i.e., "right"refers to the subject's right. Three axes are defined as x axis (left-right); y axis (anterior-posterior); z axis (inferior-superior), resulting in axial, coronal, and sagittal views of the human brain as shown in Fig. 2.4. Based on the above definitions, two conventions of the order of axes for spatial coordinates of the human brain are widely used in MRI practice: neurological convention (RAS) and radiological convention (LAS) as shown in Fig. 2.5[28]. Neurological convention is most commonly used in toolboxes of MR images

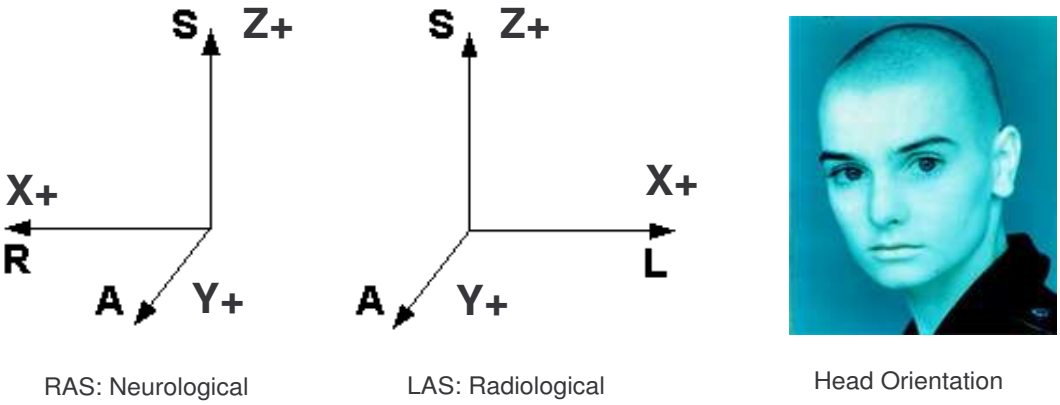


Figure 2.5: Two conventions of brain orientation.

analysis such as statistical parametric mapping (SPM), where images are viewed as though standing behind the subject and looking from the top of the head downward, i.e., the left shown on screen is left hemisphere of the brain. While in radiological convention, images are viewed as though facing the subject and looking upward from the feet of the subject, i.e., the left shown on screen is right hemisphere of the brain.

It is crucial to determine exact localizations of functional activations in the brain. In fMRI study, researchers often report a stereotaxic coordinate for functional activations of the brain. There are two coordinate systems for the reference: the atlas of Talairach and Tournoux [3] and the MNI/ICBM template developed by Montreal Neurological Institute [29]. A nonlinear transformation is developed to transfer between MNI and Talairach coordinates for results comparison and Brodmann areas (BAs) identification [30]. Besides these traditional crisp brain templates, the development of new digital atlases of the human brain is emerging, which incorporate macroscopic *in vivo* and microscopic *in vitro* data, functional information such as blood flow activation, and demographic and genetic information derived from large population by automated methods. A probabilistic digital atlas of the human brain that has been pursued since the early 1990s by the International Consortium for Brain Mapping (ICBM) [31, 32], is gaining more interests. It is derived from a large series of subjects in order to be

more representative of the entire species with information about variability. Other probabilistic atlas developed can be found in [33, 34].

2.1.4 Brain Connectivity

Early researchers in neurology and neuroscience often emphasized the localization of brain function and its distributed properties [35]. However, recent evidence from neuroimaging studies suggest that brain functions are actually dependent on the interactions among specialized brain regions that process information within local and global networks. Integration of information arises as a dynamic process on different time scales. Therefore, modern views focus extensively on the structure and dynamics of large-scale neuronal networks, especially those of the cerebral cortex and associated thalamocortical circuits underlying human perception and cognition [36, 37, 38].

A neural system is composed of numerous interconnected elements ranging from a single neuron to entire ensembles of neurons, for example, local intra-regional connections among neurons or interregional connections among ensembles of neurons across different brain areas. The human brain contains about 10 billion nerve cells. On average, each neuron is connected to other neurons via about 10000 synapses. Neurons are typically composed of a cell body, a dendritic tree, and an axon as shown in Fig. 2.6. A neuron receives input from other thousands of neurons through dendrites and transfers output to other neurons through axon. Once the input exceeds a critical value, the neuron discharges a spike - an electrical pulse that travels from the body, down the axon, to the next neuron(s). Transmission of an electrical signal from one neuron to the other is effected by neurotransmitters, chemicals which are released from the first neuron and bind to the receptors in the second. The extent to which the signal from one neuron is passed on to the next depends on many factors, e.g., the amount of neurotransmitters available, the number and arrangement of receptors,

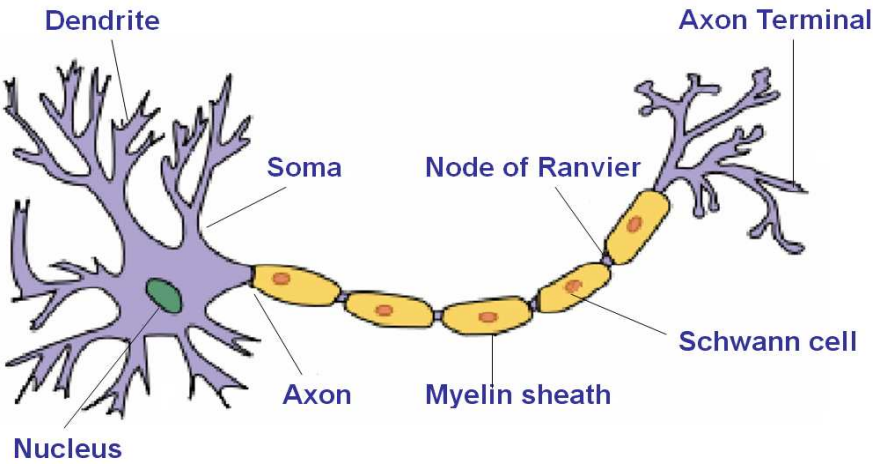


Figure 2.6: Structure of a typical neuron [39].

the amount of neurotransmitters reabsorbed and so on. Therefore, neurons maintain thousands of input and output connections with other neurons, forming a dense network of connectivity, spanning the entire thalamocortical system.

The connectivity of the human brain forms highly specific patterns, which can be interpreted at multiple levels of scale. At the lowest level, i.e., within a small and localized region of brain, neurons form characteristic sets of connections or local circuits [40]. At the higher level, local intra-area connections are responsible for specific functions of each brain region. However, characterizing brain activity in terms of functional specialization only provides a limited account of the neuronal substrates of mental processes. Inter-regional connections linking brain areas of specific sensory and motor systems form a larger-scale organization of the cortex. Recent development in the context of functional neuroimaging, therefore, aims at investigating the integration of functionally specialized areas. Three important concepts of brain connectivity are described in what follows: anatomical, functional, and effective connectivity. The interdependence and reciprocity among these three types of connectivity produce detailed insights of brain connectivity.

Anatomical connectivity refers to the set of physical or structural connections, linking

neuronal units and populations. The level of spatial scale of anatomical connectivity should be clearly identified for structural analysis, which could be local circuit level (synaptic connections between individual neurons), intra-area level (connection bundles linking local neuronal populations), and systems level (connection pathways linking segregated areas of brain) [38]. These pathways are usually comprised of many thousands or millions of individual fibers. Diffusion tensor imaging (DTI) is able to discover white matter fiber tracks, which potentially provide valuable information for anatomical connectivity of the human brain [41]. The knowledge of anatomical connectivity provides plausible biological constraints for theories and inferences of neural interactions extracted from functional neuroimaging data.

Functional connectivity is defined as the "temporal correlations between spatially remote neurophysiological events"[42]. Such temporal correlation often results from stimulus-locked transients, evoked by common afferent inputs; it may also reflect stimulus-induced phasic coupling of neuronal assemblies [43]. Functional connectivity focuses on "what the brain does", i.e., the descriptions of patterns of neural activity. Thus, functional connectivity has been reduced to testing the null hypothesis that the activity of two regions shares no mutual information. Mutual information measures the degree to which two regions demonstrate similar behavior [44]. In addition to temporal correlations, spectral coherence or consistency of relative phase relationships may also serve as indicators of functional connectivity.

Effective connectivity is defined as "the influence that one neural system exerts over another either directly or indirectly"[10]. Compared to functional connectivity, effective connectivity focuses on the explanations of the origins of brain activity or "the theory of how brain functions". The objective is to determine causal relationships among brain regions from the definitions and information of anatomical, psychological, and functional neuroimaging data. This is the major difference between functional connectivity and effective connectivity. Temporal correlations between neurophysio-

logical events in separate neural systems may or may not be due to the influence that one neural system exerts over another. There are two ways to infer effective connectivity: one is based on "temporal precedence", meaning that if activity in area A occurs prior to activity in area B, then area A could influence area B through connections between them; the other way is through perturbation studies [43].

2.2 Segmentation of Subcortical Structures

Structural magnetic resonance image provides detailed knowledge of normal and diseased anatomy for medical research and has become a critical component in clinical diagnosis and treatment planning. The segmentation of brain MR images plays a crucial role in understanding MR images by automating or facilitating the delineation of anatomical structures. It has numerous applications such as quantitative morphometry analysis, 3D volume visualization, mapping brain structures and functions, and clinical investigations including pathology, diagnosis, therapy, surgery planning, and so on [4]. Segmentation of the human brain is divided into tissue segmentation [45], subcortical structure segmentation [46], and cortical structure segmentation [47].

In our research, the focus is on the segmentation of subcortical structures in MRI. Subcortical structures of the brain are involved in quite a number of important brain functions, many of which have not been fully understood yet. Most clinical segmentation is currently performed using manual slice editing, which is time-consuming, subjective, and error prone. Inter- or intra-observer variability may reduce the detectability of subtle differences during comparisons, posing difficulties in reproducing the same results. Thus, the growing number of high resolution MR images have necessitated the use of computers to facilitate automated image segmentation [5]. However, automatic segmentation of subcortical structures is not a trivial task. Image quality and acquisition artifacts directly influence segmentation performance. Two leading

artifacts in MR images are partial volume effect and intensity inhomogeneity. Partial volume effect is referred to the mixing of different tissue types in a single voxel; intensity inhomogeneity means the spectral overlap of MR intensities of different tissue classes and the presence of a spatially and smoothly varying intensity inhomogeneity or bias field [48]. Specifically, subcortical structures often have low contrast between gray and white matter, and some gray-matter structures have striate regions which appear with intermediate intensity. In addition, the shape and size of subcortical structures vary among individuals. Other issues like spatial resolution and noise cause more challenges to segmentation.

A wide variety of methods have been proposed in the literature for segmentation of subcortical structures [5], but all of them only provide a coarse delineation of subcortical structures. Traditional low-level image processing techniques such as thresholding, region growing, edge detection, and mathematical morphology operations often require considerable amount of expert interactive guidance [4]. Low-level algorithms have to cooperate with high-level techniques such as deformable and active models or atlas-based methods to encode not just local information like intensity but also global information such as spatial and neighborhood information. Manual interaction can be used as prior knowledge to improve accuracy, but it is time-consuming, subjective, and vulnerable to reliability. An efficient way to incorporate prior knowledge is essential to segmentation. Soft segmentation based on fuzzy logic is usually used to deal with partial volume effect because of its ability to retain more information from the original image by allowing for more uncertainty in boundaries, for example, fuzzy c-means classifiers. In addition, validation experiments are necessary to quantify the performance of a segmentation method. One direct approach is to compare automated segmentations with manually obtained segmentations. The other approach is to use physical phantoms or simulated phantoms. A survey about segmentation validation is available in Zhang et al. [49].

In the following, three main techniques in the literature for segmentation of subcortical structures are presented: deformable models, atlas-based models, and hybrid models.

Deformable models, which stems primarily from the theory of elasticity at the physical level, is a promising and vigorously researched model-based approach to computer-assisted medical image analysis, including image segmentation. Various names have been used to refer to deformable models such as snakes, active contours or surfaces, balloons, and deformable contours or surfaces. The mathematical foundations of deformable models represent the confluence of geometry, physics, and approximation theory. Geometry serves to represent object shape, physics imposes constraints on how the shape may vary over space and time, and optimal approximation theory provides the formal underpinnings of mechanism to fit the models to measured data [4].

Typically, in the case of segmentation of subcortical structures, a deformable model delineates region boundaries using curves or surfaces that deform under the influence of internal and external forces. Internal forces keep the shape smooth throughout the deformation and external forces drive the boundaries to the desired location. A deformation energy function defined in terms of these forces is associated with the prior deformable model and is minimized to achieve optimal segmentation. Generally, deformable contours (2D) and deformable surface (3D) of structures are physically motivated models. There are two kinds of deformable models: parametric and geometric. Parametric models introduced in Kass et al. are widely employed with encouraging results, and they represent curves and surfaces of the structures explicitly by parameters that lead to direct interaction and fast real-time implementation [50]. Geometric models are based on curve evolution and level set method, which can handle topological changes of structures more naturally than parametric models [51].

The main advantage of deformable models is their ability to directly generate closed

curves or surfaces from images and their smoothness constraints that provide robustness to noise and spurious edges [5]. Since most anatomical structures are deformable and continually undergo nonrigid motion *in vivo*, a dynamic deformable model is constructed to measure shape evolution through time. Prior knowledge could be easily incorporated as well. However, deformable models rely on human experts for initialization and guidance, and intelligent optimization algorithms are required to automate the approach. Recently, fuzzy logic, level set theory, and genetic algorithm are incorporated into deformable models to improve the accuracy of segmentation of subcortical structures [52, 53, 54].

Atlas-based approaches have been applied to segmentation of brain structures, extraction of brain volume, and measurement of brain morphometry from MRI [55, 56, 57]. Since prior knowledge is very helpful to segmentation of brain subcortical structures, atlas-based approaches provide a powerful tool if a standard atlas or template is available. Elastic image registration is used to map a pre-segmented atlas to the target image to compute linear or nonlinear transformation [5]. This process is often called atlas warping. Atlas-based approaches are better for segmenting structures that are consistent across a population of subjects. However, inter-subject variability indeed exists in brain subcortical structures. In order to address the problem of anatomical variability, probabilistic atlas is proposed [31, 32]. Fischl presented a technique for assigning a neuroanatomical label to each voxel in a MRI volume based on probabilistic atlas that is automatically estimated from a manually labelled training set [58]. The classification procedure includes atlas construction, optimal linear transform that aligns each subject image anatomically, segmentation based on ICM (iterated conditional modes) algorithm, and Bayesian models to compute the MAP estimate. Collins also used a probabilistic anatomical atlas and a non-linear registration technique to produce gross anatomical region mask [47]. Nevertheless, the accuracy of registration mechanism largely influences the segmentation performance of atlas-based approaches.

On the other hand, the accuracy of atlas-based registration often degrades when the data are acquired on different scanner platforms or pulse sequences from the data used in training. An intensity re-normalization procedure was introduced to automatically adopt the intensity model of prior atlas to the new input images[57].

Recently, however, segmentation approaches are moving towards knowledge-based hybrid models where the advantages of region-based, edge-based, deformable-models, and atlas-based methods are exploited for the integration or fusion of information in a systematic way. Examples are fuzzy modelling [59], information fusion [46, 60], the combination of fuzzy clustering with deformable models [61], integration of fuzzy spatial relations in deformable models [62], and so on. Other approaches include histogram analysis [63], Bayesian approach [58], neural network techniques [64, 65], and genetic algorithms [66]. An important aspect of this research is to extract prior knowledge of anatomical features from training data and use them as constraints or markers for deformation or registration [67, 57, 68]. Despite all these efforts, the accuracy of segmentation of brain subcortical structures from MR images is inadequate for clinical practice. Moreover, novel methods should reduce computational speed as well as the amount of manual interactions.

2.3 Detection of Activation from fMRI

Based on blood-oxygenation-level-dependent (BOLD) signal, functional magnetic resonance imaging (fMRI) is able to reveal unprecedented insights into spatial and temporal changes underlying a broad range of brain functions and has enhanced our understanding of the neural basis of human behavior. Raw data from a fMRI experiment is analyzed with the aim to accurately determine the activated regions of the brain, where the signal changes upon the presentation of stimuli. However, fMRI signal changes are subtle, for example, on a 1.5 T scanner, the BOLD signal changes

of the brain due to the experimental stimuli are approximately $1\% - 5\%$ [19]. The variability of hemodynamic response function (HRF) across brain regions and individuals is another important factor for consideration. In addition, various noise and artifacts such as motion and physiological processes significantly confound the fMRI signal. Therefore, accurate detection of functional activation from fMRI data should be insensitive to the interfering signals [69].

The process of detection of brain activation from fMRI data consists of three phases: (1) pre-processing such as realignment for movement correction, co-registration with structural images (MRI), normalization to a standard template for easy comparison, and smoothing for signal to noise ratio enhancement (more details is described in section 2.3.1); (2) detection of activated voxels or regions by hypothesis-driven or data-driven methods; and (3) the labelling of activation for identification and detailed visualization of brain regions. In activation labelling process, the detected activation could be labelled in terms of stereotaxic coordinates, macroanatomy, function, or microanatomy. For some brain areas, all of these attributes are inter-related, which makes the characterization straightforward. In many other brain areas, the correspondence is yet more difficult to determine. Therefore, it is crucial to choose the correct activation labelling technique for visualization and comparison among different groups of subjects.

There are mainly two groups of approaches for detection of activation from fMRI data: hypothesis-driven and data-driven. Hypothesis-driven methods assume a specific model for the fMRI signal with a specific noise structure while data-driven methods do not require *a priori* knowledge about patterns of signals. The techniques of pre-processing and two groups of methods for the detection of activation from fMRI are explained in the following sections.

2.3.1 Preprocessing of fMRI Data

For fMRI time-series analysis, structural and functional MRI are usually collected together to identify functionally activated anatomical brain regions. Therefore, preprocessing, i.e., the analysis of structural and functional data from different scans of the same subject, or data from different subjects, is a necessary step before statistical analysis to make images conform to the same anatomical frame of reference. Main steps include realignment and unwarp, co-registration, normalization, and smoothing. Note that realignment, co-registration, and normalization are all fundamental problems in medical imaging, for which many different techniques have been attempted. Detailed review of these methods is out of the scope of this thesis. The methods used in our experiments as implemented in statistical parametric mapping (SPM), the most widely used software of fMRI analysis, are presented here.

1. *Realignment and Unwarp*

   Intra-modality and intra-subject image registration is useful for correction of motion artifacts in fMRI time-series. Under the assumption of no shape change and rigid body motion, registration is achieved by optimizing the parameters describing a rigid body transformation between the source and reference images and then re-sampling according to the determined transformation [70]. The 3D rigid body transformation is denoted by translations in X, Y, and Z coordinate

directions and rotations about X, Y, and Z axes:

$$\begin{pmatrix} 1 & 0 & 0 & X_{trans} \\ 0 & 1 & 0 & Y_{trans} \\ 0 & 0 & 1 & Z_{trans} \\ 0 & 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\Phi & \sin\Phi & 0 \\ 0 & -\sin\Phi & \cos\Phi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} \cos\Theta & 0 & \sin\Theta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\Theta & 0 & \cos\Theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} \cos\Omega & \sin\Omega & 0 & 0 \\ -\sin\Omega & \cos\Omega & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.1}$$

where X, Y, and Z axes are defined as in section 2.1.3.

An optimization procedure is required to find the best parameters according to suitable objective function. In this case of intra-modality, intra-subject realignment, a series of 3D images are produced by the same MR machine of the same subject in one session so voxel intensity relationship is assumed to be normally distributed. Gauss-Newton algorithm is proposed to iteratively optimize the transformations by minimizing mean-square differences between a series of images and a reference image [70]. The reference image can be one of the series or the mean of all images for slightly better results. After all the transformations are determined, images should be re-sampled to perform the transformations. Interpolation is used for re-sampling such as nearest neighbor interpolation, trilinear interpolation (a weighted average of neighbors), and B-spline interpolation (default option in SPM). More complex interpolation results in longer computational time.

A more robust solution may make use of the symmetrical property of registration problem, which means registering from image A to image B should be simultaneously optimized with the inverse registration, so the cost function should

include the differences in both directions [71, 72]. Moreover, a variance image is computed for optional registration weighting due to non-stationary variance in images, meaning that voxels with high variance should have a lower weight. After realignment, there are still some motion related artifacts that cannot be solved by rigid transformations [73, 74]. Furthermore, susceptibility-by-movement artifacts exist in fMRI time series because material has its own susceptibility leading to different field disturbance caused by the presence of different objects. Unwarp is required in addition to realignment to address this problem [75].

2. *Co-registration*

In a fMRI study, this intra-subject and inter-modality image registration is an essential step to co-register a series of functional MR images of a particular subject to the corresponding structural MR image. Because only when structural and functional images are co-registered, the high-resolution structural MR image can be superimposed on fMRI to provide detailed anatomical background for functional activation maps of the brain. The co-registration also enables the study of the relationship between brain structures and functions of individuals.

An elegant approach is proposed to match images by maximizing the mutual information of their histograms, which has been applied to many image registration problems among different modalities. Generally, the registration involves maximizing a cost function to compute the optimal transformation. Several cost functions are applicable including mutual information [76, 77], entropy correlation coefficient [78], and normalized mutual information [79]. A good survey of this method could be found in [80, 81]. Another co-registration approach relies on prior probability images of gray matter (GM), white matter (WM), and cerebro-spinal fluid (CSF) that conform to the standard template image [82]. It consists of three steps: normalize different modality images to its own

template, segment the images using probability images and a modified mixture model algorithm, and co-register the image partitions by optimizing the rigid body transformation between partitions of each modality.

3. *Normalization*

Normalization, a within modality and between subject spatial transformation, matches an arbitrary image to some ideal image, model, or template such that homologous regions from different subject are as close together as possible. Normalization to a standard template is necessary in neuroimaging analysis, especially for interpretation and comparison of group studies. Firstly, the inter-subject averaging of functional activation requires converting different images to the same space to increase sensitivity. A further advantage is that activation sites are reported according to their Euclidian coordinates within a standard space [83].

There are several issues in this normalization step. Firstly, it is important to know how anatomy is related to function. In other words, we need to know the basic assumptions of the relationship between anatomy and function of the brain and evaluate whether a particular spatial normalization method automatically results in good functional normalization. Inter-subject variability in brain anatomy produces more uncertainty in normalization. On the other hand, the choice of atlas is of practical importance. Traditionally, Talairach atlas coordinate system, a 3D proportional grid system with reference to anterior commissure (AC) and posterior commissure (PC) points, has been widely used [3]. Nevertheless, because it is based on a single subject, Talairach atlas cannot represent the common characteristics of a population. Montreal Neurological Institute (MNI) has created a standard template by averaging a large number of MRI scans of healthy young adults, which is more representative of a population. Since most of the previous work in fMRI study reports activation in

Talairach coordinates, a nonlinear transformation is often required to transfer MNI coordinates to Talairach coordinates for comparing results and estimating Brodmann areas (BAs) [30].

Currently, there are two normalization methods: volume-matching and sulcal-matching. A more intelligent approach is needed to match both cortical and subcortical brain areas [2]. Normalization in SPM uses a maximum a posteriori (MAP) solution, which first determines the optimum 12 parameters affine transformation by MAP optimization algorithm [70] with prior knowledge of variability of head size [84]. Secondly, a nonlinear transformation for correcting gross differences in head shapes is modelled by a linear combination of smooth basis functions. The parameters of this combination in nonlinear transformation are optimized by a fast algorithm [85]. Furthermore, if the co-registration between functional and structural images exists, a precise spatial normalization can be achieved by computing mapping from high-resolution structural MRI to the T1-weighted template. The resulting spatial transformation together with the co-registration mapping between structural and functional images can be applied to fMRI time-series at the same time and re-sampling is performed only once for these two steps.

4. *Smoothing*

Besides realignment, co-registration, and normalization, smoothing is often required at last to increase signal to noise ratio before entering the next detection phase. In SPM, an isotropic Gaussian kernel with a specified width is used and smoothing is a necessary step to make sure that images have the characteristics of a Gaussian random field (GRF) to satisfy the statistical assumptions of SPM. Since biological models usually imply contiguous activation areas, i.e., neighboring voxels have similar activation status, spatial smoothing is able to apply spatial regularization to remove potential small false positive islands.

Other preprocessing tools for fMRI analysis are slice timing to adjust timing differences in multi-slice image acquisition and tissue segmentation. In summary, preprocessing of fMRI data is essential for the proper characterization of evoked hemodynamic changes in fMRI, in terms of both artifacts removal and anatomical localization.

2.3.2 Correlation-based Methods

Correlation methods, one of the early hypothesis-driven techniques, was first applied to fMRI data by Bandettini et al. in 1993. Assume the time-series of a block-designed fMRI experiment is properly realigned, a reference function is created by convolving the block design stimulus with a fixed hemodynamic response function (HRF). The modulated stimulus is then correlated with fMRI time-series of each voxel in the image to determine the areas of activation [86]. The correlation coefficient r_i, also known as the Pearson's correlation coefficient, between the recorded time-series $y_i = \{y_{it} : t = 1, 2, ...T\}$ of i^{th} voxel with the modulated reference input x is given by:

$$r_i = \frac{\sum_{t=1}^{T}(y_{it} - \bar{y}_i)(x_t - \bar{x})}{\sqrt{\sum_{t=1}^{T}(y_{it} - \bar{y}_i)^2 \sum_{t=1}^{T}(x_t - \bar{x})^2}} \tag{2.2}$$

where t refers to time point. Voxels having high correlation coefficient are related positively to the reference function and are designated as the areas of activation. A correlation of 0.4 is typically used as the threshold value.

2.3.3 Statistical Parametric Mapping (SPM)

Statistical parametric mapping (SPM), the most widely used hypothesis-driven approach, refers to the conjoint use of the general linear model (GLM) and Gaussian random field (GRF) theory to analyze and make classical inferences about spatially

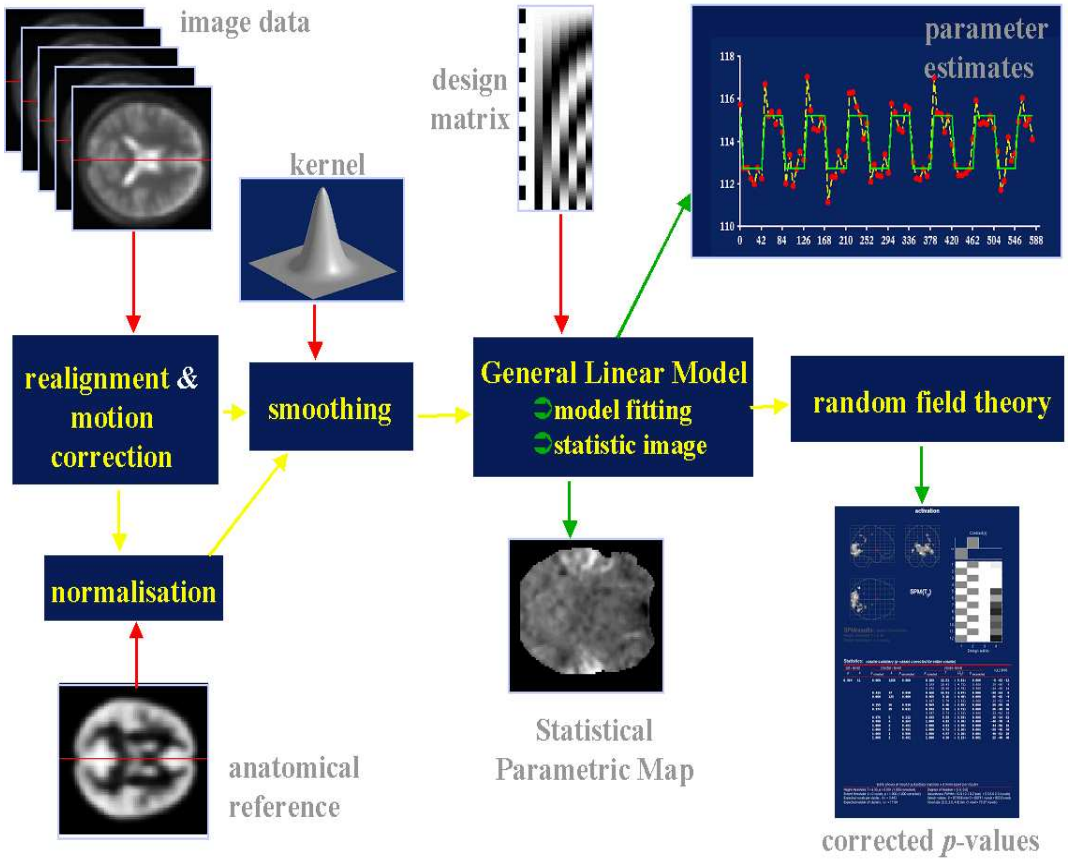
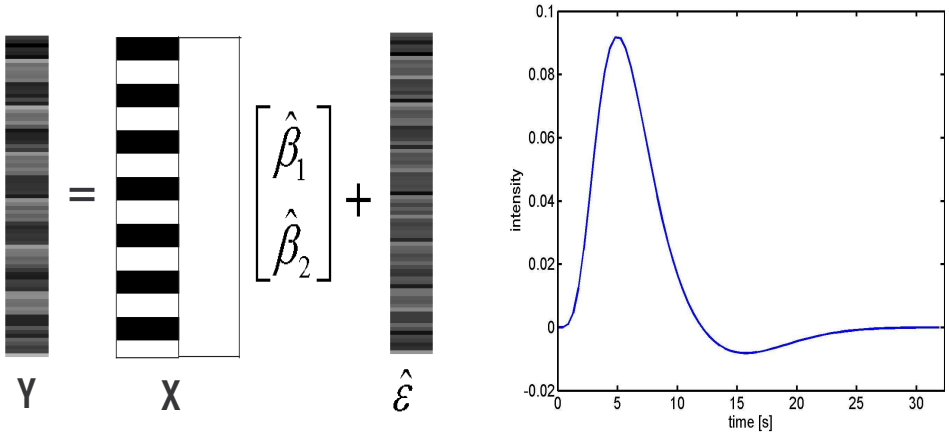


Figure 2.7: Flowchart of Statistical Parametric Mapping [89].

extended data through statistical parametric maps [87, 88]. SPM is a voxel-based method assuming a linear regression model for the fMR signal with a specific noise structure. These ideas have been instantiated in a software referred as SPM by Wellcome Department of Imaging Neuroscience [89]. SPM is composed of two main steps for fMRI analysis: preprocessing and statistical analysis, as illustrated in Fig. 2.7. The former step is described in section 2.3.1 while the latter step, statistical parametric mapping for fMRI, is performed as follows.

The objective of general linear model (GLM) in SPM is to partition data into effects (of interest and no interest) and error, and to find the optimal value of the parameter β of the model to make statistical inferences. The GLM can be seen as multiple



(a) General Linear Model (b) Hemodynamic response function

Figure 2.8: Illustration of general linear model (GLM) and hemodynamic response function (HRF) of the brain.

regression, which is written in matrix notation for a time-series as [90]:

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\beta + \mathbf{e} \\ &= \mathbf{H}\eta + \mathbf{D}\gamma + \mathbf{e} \end{aligned} \tag{2.3}$$

where $\mathbf{Y}$ is a column vector of observed time-series from a single-voxel; $\mathbf{X} = [\mathbf{HD}]$ is the design matrix composing a set of basis functions, which has one row for every scan and one column for every effect (factor or covariant) in the model; the column of $\mathbf{H}$ model the effects of interest, e.g., one reference waveform; the columns of $\mathbf{D}$ model confounds - effects of no interest, e.g., global activity of the scan; β is a column vector of parameters with one row for each effect modelled by each column of design matrix; the error $\mathbf{e}$ is assumed to be independent and identically normally distributed with mean zero and covariance σ^2. This is illustrated in Fig. 2.8(a).

SPM is based on the assumption that interesting hemodynamics are the result of convolving an underlying neuronal process with a hemodynamic response function (HRF) [90]. More specifically, it is presumed that fMRI time-series has a number of linearly separable components that including high frequency noise, low frequency noise, and

low frequency signal. The relative amounts of these components change with the field strength, pulse sequence, and experimental design. The signal of interests is always smoother than the response function. Therefore, temporal smoothing of fMRI data always increases interesting hemodynamic variance components relative to other components, hence, increases signal to noise ratio at the cost of a reduction in the effective degrees of freedom. However, temporal smoothing renders scans correlated in time, which violates the fundamental assumption of the original general linear model - each scan is an independent observation. The basic general linear model is then extended by including temporal smoothing so that it can be applied to data with a stationary and known autocorrelation [7].

Let $\mathbf{K}$ be a convolution matrix applied to original data $\mathbf{Y}$. The optimal kernel for smoothing is related to the hemodynamic response function (HRF) of the brain as shown in Fig. 2.8 (b), which is approximated by a Gaussian kernel with parameters representing dispersion and delay of the HRF. Then Eq. 2.3 becomes

$$\mathbf{KY} = \mathbf{X}^*\beta + \mathbf{Ke} \tag{2.4}$$

where $\mathbf{KY}$ represents temporally smoothed data and $\mathbf{X}^* = \mathbf{KX}$. Error term is identically distributed with covariance $\sigma^2\mathbf{KK}^T$. The optimal value of β for this model maximizes variance in the signal frequencies relative to other frequencies, which is achieved by minimizing least squares of error as:

$$\hat{\beta} = (\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}\mathbf{KY} \tag{2.5}$$

where $\mathbf{X}^* = \mathbf{KX}$, $E\{\hat{\beta}\} = \beta$, and variance of $\hat{\beta}$ is given by

$$\mathrm{Var}\{\hat{\beta}\} = \sigma^2(\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}\mathbf{V}\mathbf{X}^*(\mathbf{X}^{*T}\mathbf{X}^*)^{-1} \tag{2.6}$$

where the unbiased estimator $\hat{\sigma}^2$ of the real error variance term σ^2 is obtained by dividing the residual sum of squares by its expectation [7].

Based on the statistical model of estimated parameters $\hat{\beta}$ fitted on fMRI data, the effect of interests, e.g., the difference between a task and baseline, can be investigated. Statistical inferences are performed by testing the null hypothesis that the effect embodied in $\mathbf{H}$ is not significant for a specific brain voxel. To do this, different effects of interests are represented with a linear compounds or contrasts of $\hat{\beta}$ and the classical T statistic is used:

$$T = \frac{c^T \hat{\beta}}{\sqrt{\mathrm{Var}(c^T \hat{\beta})}} \tag{2.7}$$

where contrast vector c has the same length as $\hat{\beta}$, and $\mathrm{Var}(c^T \hat{\beta})$ is given by:

$$\mathrm{Var}(c^T \hat{\beta}) = \hat{\sigma}^2 c^T \cdot \mathrm{Var}\{\hat{\beta}\} \cdot c \tag{2.8}$$

If the computed T for the effect of interest in one voxel is larger than a specific threshold T^* at a certain confidence level, the null hypothesis is rejected and this voxel is determined to be activated. To test $K > 1$ number of contrasts at the same time, the T statistic is replaced by F statistic defined as in Eq. 2.9. For simplicity, the T statistics is transformed to Z score using a probability integral transform or other standard device, and we try to interpret the statistical parametric map of Z score in terms of probability levels or p-values.

$$F = \frac{\hat{\beta}^T c \cdot (\hat{\sigma}^2 c^T \cdot \mathrm{Var}\{\hat{\beta}\} \cdot c)^T c^T \hat{\beta}}{K} \tag{2.9}$$

However, as there are many thousands of voxels in the brain, a large number of non-independent univariate comparisons need to be performed. The probability that any voxel exceed an uncorrected threshold by chance is high in this multiple comparison case [90]. Therefore, a method for threshold correction is required. Because of the

spatial correlations in statistical parametric maps, the usual solution - Bonferroni correction is not suitable. Random field theory (RFT) is a recent branch of mathematics about smooth statistical maps used to solve this problem for statistical inference [91]. Details of underlying theory are not reviewed here and are found in [92, 93]. RFT is applied to correct for the multiplicity of voxels and the spatial correlations among them in SPM. The application of RFT requires two assumptions: (a) a statistical image is a good lattice approximation to underlying multivariate Gaussian fields; (b) these fields are continuous with a twice differentiable correlation function. RFT works by calculating the expected Euler characteristic (EC) for a smooth statistical map that has been thresholded, which in turn gives the optimal threshold that we need. The procedure is briefly summarized as follows [91]:

- *Estimate the smoothness*: The spatial correlation or smoothness of statistical map is expressed as the width of the smoothing kernel, for example, full width at half maximum (FWHM) kernel. Let r be the n-dimensional vector of least-squares residuals from the linear model fitted at each voxel, u be the vector of normalized residuals $u = r/(r'r)^{1/2}$, and $\dot{u}$ be the $n \times 3$ spatial derivative of u in the three orthogonal directions of the voxel lattice. The FWHM kernel is estimated using the residual values of statistical maps as:

$$\widehat{FWHM} = (4log2)^{1/2}|\dot{u}'\dot{u}|^{-1/(2D)} \tag{2.10}$$

  where D is image dimension.

  A resel is "resolution element", defined as a block of voxels that have the same size as the FWHM. The number of resels can be thought of as similar to the number of independent observations in the image, which is useful for Euler characteristics (EC) calculation in the next step. The estimation of resel count

is given by:

$$\widehat{Resels}_D = \sum_{volume} \widehat{FWHM}^{-D} \cdot v \tag{2.11}$$

where "volume"means that summations is over all voxels in the search region and v is the volume of a single voxel [94].

- *Produce the expected Euler characteristics (EC)*: The EC is a property of an image after it has been thresholded, which can be considered as number of blobs in the image. The expected EC of the image at different thresholds is calculated by using the smoothness values in terms of resel count as:

$$\overline{EC} = \widehat{Resels}_D \cdot (4 \log_e 2)(2\pi)^{-\frac{3}{2}} Z_t e^{-\frac{1}{2}Z_t^2} \tag{2.12}$$

where Z_t stands for the Z score threshold.

- *Calculate the threshold*: One useful feature of the expected EC is that it drops towards zero when the Z score threshold becomes very high. Therefore, the expected EC gives the correct threshold for the required control of false positives.

The above threshold correction procedure based on RFT can be considered as Bonferroni correction on resels, in which familywise null hypothesis is tested to control the familywise error rate(FWER), i.e., the probability of reporting a false positive anywhere in the volume [95, 96]. False discovery rate (FDR) is another approach to solve the problem of multiple comparisons. Instead of controlling the chance of any false positives, FDR controls the expected proportion of false positives among rejected (positive) voxels [97, 98]. A FDR threshold is determined from the observed p-value distribution, and hence is adaptive to the amount of signal in fMRI data. FDR is more sensitive than traditional methods because of its more lenient metric for false positives. Finally, after determining the optimal threshold, the voxels having Z score higher than the threshold are considered as activated.

2.3.4 Data-driven Methods

Data-driven methods do not rely on any assumed model of functional response and are considered more powerful and relevant for fMRI analysis [6], especially when activation pattern is unknown or complex. There are generally two types of data-driven methods: transformation-based and clustering-based.

Principle component analysis (PCA) [99] and independent component analysis (ICA) belong to the transformation-based methods, which transform the original data into high-dimensional vector space to separate functional responses from various noise sources and artifacts. A major problem of PCA is that it accounts for only second-order dependencies among components. ICA is an information theoretic approach which enables recovery of underlying signals or independent components from linear data mixtures. It has been applied to fMRI analysis both spatially [100] and temporally [101, 102]. ICA reduces higher order dependency, but is still limited by assumptions of stationary distribution and linear data mixtures. Recently, constrained ICA [103] or the fusion of temporal and spatial ICA [104] has been developed to increase detection accuracy and to provide more meaningful components. On the other hand, the clustering approaches include fuzzy clustering [105] and self-organizing map [9], attempting to classify time signals of the brain into several patterns according to temporal similarity. They both have difficulty in finding the activation in a small area and they promote sensitivity to noise and initial values.

Recently, researchers have modelled spatial coherency in the detection of activation from fMRI using Gaussian smoothing [89], Markov random fields (MRF) [106], and conditional random fields (CRF) [107]. Such techniques account for contextual dependencies among activated status or data of neighboring voxels and has the potential to remove small false positive islands. Other fMRI analysis methods include specified-resolution wavelet analysis [69, 108], and Bayesian modelling [109].

In summary, hypothesis-driven methods assume a specific model for the signal and noise in fMRI data. They are voxel-based with potential contextual modelling. However, it has become clear that there is a nonlinear relationship between the variations in the fMRI signal and the stimuli [110]; and the hemodynamic response function (HRF) varies spatially and across subjects [109]. The structure of noise in fMRI is not well understood and remains a contentious subject. The validity of the statistical models depends on the extent to which the data satisfies the underlying assumptions. Thus hypothesis-based methods have some limitations on fMRI analysis when complicated experimental tasks are used or no prior knowledge of activation patterns are available. In contrast, data-driven methods detect activation in an exploratory manner without any assumptions of signals or noise, and thus are more suitable for fMRI studies of complex or unknown activation patterns. Nevertheless, it is difficult to ascertain how the fMRI data is divided into classes. Certain class with a particular activation pattern is physiologically interpreted as activation, but the interpretations of many other classes or components are still unknown. Besides the above mentioned issues, different block or event-related experimental designs further complicate the underlaying physiological processes, adding more challenges to accurate detection of brain activation from fMRI data.

2.4 Inference of Effective Connectivity

Explanation of functional specialization from neuroimaging data does not provide information of how brain regions communicate with each other in order to materialize a particular task. Functional integration studies in terms of functional and effective connectivity aim at addressing the regional neuronal interactions to obtain a holistic view of brain functions. As effective connectivity studies the mechanisms of brain functions rather than functional connectivity, which only provides the description of

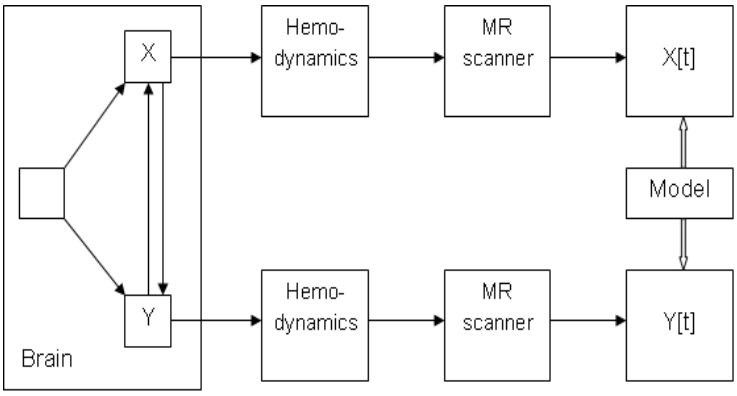


Figure 2.9: Illustration of indirect access to interacting brain regions with fMRI. Hemodynamics and the MR scanner contribute unwanted artifacts to the signals of interest and might confound modelling efforts. Confounding is especially deleterious when the unwanted contributions are different for the brain regions under investigation (e.g., different hemodynamic responses in different regions) [14].

brain functions, we focus on learning effective brain connectivity from fMRI data in this thesis.

There are several issues related to learning effective brain connectivity from fMRI data. Some methods like GCM assume that fMRI time-series contain sufficient temporal information to determine directed influence based solely on temporal precedence. Though the sub-second range of temporal resolution in fMRI may not be sufficient to reveal interactions among individual neurons, fMRI still provides temporal dependencies among cognitive components that are useful for interpreting complex tasks [111]. An inherent problem in inferring interactions at the level of neuronal populations from fMRI data is the indirect access to signals of interest (Fig. 2.9). FMRI signal can be considered as a filtered and sampled version of the local field potential (LFP) signal rendering a measure of the fluctuations of activations of local neuronal populations [21]. Additionally, the differences in the hemodynamic response function (HRF) among brain regions across different subjects and cognitive tasks bring more difficulty to learning effective brain connectivity.

Variability of effective brain connectivity exist due to: (1) intra-subject variability,

Method	Details	Pros and Cons
SEM	Decompose interregional covariances of activations	Second-order statistics; no causal effects are reflected
DCM	A bilinear model for neurodynamics and an extended balloon model for hemodynamics	Model at the neuronal level rather than hemodynamic level; need a prior model to begin with
GCM	Use a vector autoregressive (VAR) model based on Granger causality	Linearity; causes of fluctuations are not explicitly modelled; not applicable to large networks

Table 2.2: An overview of three techniques for brain connectivity analysis.

i.e., different runs of the same subject could have different effective connectivity for a particular task; (2) inter-subject variability, i.e., effective connectivity for a same task could vary among individuals; (3) variability across different cognitive tasks; and (4) the time-variant nature of connectivity. It is therefore important to address the variability of effective connectivity of the brain.

This section reviews the earlier approaches in inferring brain connectivity from fMRI data: structural equation modelling (SEM), dynamic causal modelling (DCM), and Granger causality mapping (GCM). Table 2.2 presents an overview of these three methods.

2.4.1 Structural Equation Modelling (SEM)

Structural equation modelling (SEM) was developed in the field of econometrics and first applied to imaging data by McIntosh and Gonzalez-Lima [11]. Later, SEM was applied for network analysis of vision tasks using PET [112, 113]. Other researchers applied SEM for the analysis of networks of brain regions involved in sensory [114], language [115, 116, 117, 118], and cognitive tasks [119, 120] from PET, EEG, and fMRI data. To date, SEM is the most commonly used method to analyze brain connectivity from neuroimaging data, which extracts the information about neural

interactions by decomposing interregional covariances among activations. For a certain cognitive task, SEM is applied to confirm or reject the hypothesis of a known structure. Moreover, Bullmore *et al.* showed how to search for the best fitting covariance model of connectivity from fMRI data by using SEM [13]. Mechelli *et al.* [121] constructed multi-subject networks based on SEM to illustrate the differences in connectivity among subjects.

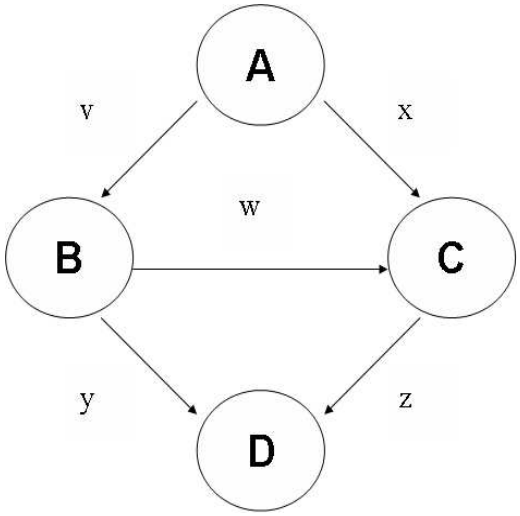
Fig. 2.10 presents a prior model of a set of variables. The covariance structure among these variables are described by SEM and the connections in the structural model imply a particular set of instantaneous correlations among variables. The strength of connections are set to minimize the discrepancy between the observed and predicted correlations and thereby find the best model fitting the data [122]. The fundamental hypothesis of the structural equation procedure is that the covariance matrix of the observed variables is a function of a set of parameters. Let Σ be the population covariance matrix of observed variables and $\Sigma(\theta)$ represents the covariance matrix predicted by the model, where θ is a vector that contains the model parameters.

In the context of inferring connectivity from fMRI data, the connections in a theoretically anticipated prior path model among n brain regions are represented by a matrix, and the estimated path coefficients are computed by minimizing a measure of discrepancy between the observed interregional correlation matrix C and the correlation matrix predicted by the model $\Sigma(\theta)$ [13]. The pairwise correlations c_{ij} between the first eigen time-series of the ith and jth regions constitute the interregional correlation matrix C. The residual variance of each region is estimated by the ratio between the first eigenvalue and the sum of eigenvalues as:

$$\Psi_i = 1 - \frac{{\lambda_1}^2}{\sum_{j=1}^{m} {\lambda_j}^2} \tag{2.13}$$

where λ_i denotes the ith eigenvalue of the correlation matrix.

(a) Structural Model



(b) Correlation Matrix

	A	B	C	D
A	---			
B	r_{AB}	---		
C	r_{AC}	r_{BC}	---	
D	r_{AD}	r_{BD}	r_{CD}	---

(c) Path Equations

$$r_{AB} = v$$

$$r_{AC} = x + (vw)$$

$$r_{AD} = (vy) + (xz) + (vwz)$$

$$r_{BC} = w + (vx)$$

$$r_{BD} = y + (wz) + (vxz)$$

$$r_{CD} = z + (wy) + (wvy)$$

(d) Structural Equations

$$A = \psi_A$$

$$B = \nu_A + \psi_B$$

$$C = x_A + w_B + \psi_C$$

$$D = y_B + z_C + \psi_D$$

Figure 2.10: Schematic representation of methods involved in structural equation modelling of a neural system: (a) path diagram of a simple network with four regions; (b) the information about the correlations of activities; (c) path equations showing how the correlations between regions can be decomposed to solve for the path coefficients; and (d) structural equations showing the variance in activity of each region as a function of the weighted variance of other brain regions and a residual influence. [113]

The SEM model is written down as a set of simultaneous regression equations:

$$x = Kx + \psi \tag{2.14}$$

where x denotes the vector of regional variances, ψ denotes the vector of residual variances, and K denotes the matrix of the path model. Eq. 2.14 can be rearranged as:

$$x = (1 - K)^{-1}\psi$$

and the correlation matrix $\Sigma(\theta)$ predicted by the path model is given by:

$$\Sigma(\theta) = (1 - K)^{-1}\psi\psi'(1 - K)^{-T}$$

Thus, the estimates of the path coefficients are found by iteratively minimizing the maximum likelihood (ML) discrepancy function:

$$F = \log|\Sigma(\theta)| + \mathtt{tr}(C\Sigma^{-1}(\theta)) - \log|C| - q \tag{2.15}$$

where $|.|$ denotes the determinant and $\mathtt{tr}$ denotes the trace of a matrix.

Although SEM has been widely applied to infer brain connectivity in neuroimaging community, there are still some issues related to it. The covariances between the brain regions in SEM limit the behavior of a neural system to second-order statistical sense. It only renders functional connectivity and fails to establish sufficient conditions for the identification of SEMs with loops (reciprocal connections), which are biologically common models [123].

2.4.2 Dynamic Causal Modelling (DCM)

Dynamic causal modelling (DCM) was first introduced by Friston [12] to estimate and make inferences about the coupling among brain regions and how such coupling is influenced by the changes of the experimental context. Coupling here refers to interactions among neuronal populations at a cortical level based on their hemodynamic or electromagnetic properties. Conventional SPM for fMRI analysis discounts the coupling among brain regions. DCM potentially extends the GLM model for making inferences about the temporal changes of effective connectivity from fMRI data [12, 123]. The tenet of DCM is the construction of a reasonably realistic neuronal model of interacting cortical regions, which is neurophysiologically meaningful with parameters quantifying the brain functions [12]. Dynamic causal modelling is specially designed for fMRI data to model functional interactions at the neuronal level, which comprises of a bilinear model for neurodynamics and an extended balloon model for hemodynamics. The parameters are estimated such that the predicted BOLD time-series, resulting from the conversion from neural dynamics into hemodynamics, corresponds as closely as possible to the observed BOLD time-series.

Let us denote neuronal activity by z and the inputs by u. Dynamic causal modelling approach models the dynamics of neural states $\dot{z}$ as a non-linear function of the states as

$$\dot{z} = F(z, u, \theta) \tag{2.16}$$

A bilinear model is proposed to approximate the above nonlinear model (Eq. 2.16) in

a natural way by reparameterisation in terms of effective connectivity:

$$\hat{z} \approx Az + \sum u_j B^j z + Cu = (A + \sum u_j B^j)z + Cu \tag{2.17}$$

$$A = \frac{\partial F}{\partial z} = \frac{\partial \hat{z}}{\partial z} \tag{2.18}$$

$$B^j = \frac{\partial^2 F}{\partial x \partial u_j} = \frac{\partial}{\partial u_j} \cdot \frac{\partial \hat{z}}{\partial x} \tag{2.19}$$

$$C = \frac{\partial F}{\partial u} \tag{2.20}$$

where matrix A represents the effective connectivity among brain regions in the absence of modulatory input, matrices B^j encode the changes in effective connectivity induced by the jth input u_j, and C embodies the strength of direct influences of inputs on neuronal activity.

Fig. 2.11 shows the forward model of hemodynamics, translating neuronal activity into a BOLD response. The parameters and time constants of the neuronal model is estimated from measured data by a Bayesian approach using empirical priors for the biophysical parameters and conservative shrinkage priors for the coupling parameters. The inference is done by testing hypothesis using the posterior distributions of the parameter estimates. If there is uncertainty about which connections should be included in a model, or if one would like to compare competing hypotheses (represented by different DCMs), Bayesian model selection procedure is used to determine the DCM that shows an optimal balance between model fit and the number of parameters. More recently, Penny *et al.* [124] proposed a bilinear dynamical systems (BDS) for model-based deconvolution of fMRI time series, which comprises of a stochastic bilinear neurodynamical model specified in discrete-time and a set of linear convolution kernels for hemodynamics. DCM is not an exploratory technique and the results are specific to the prior models used for the tasks and stimuli in fMRI study.

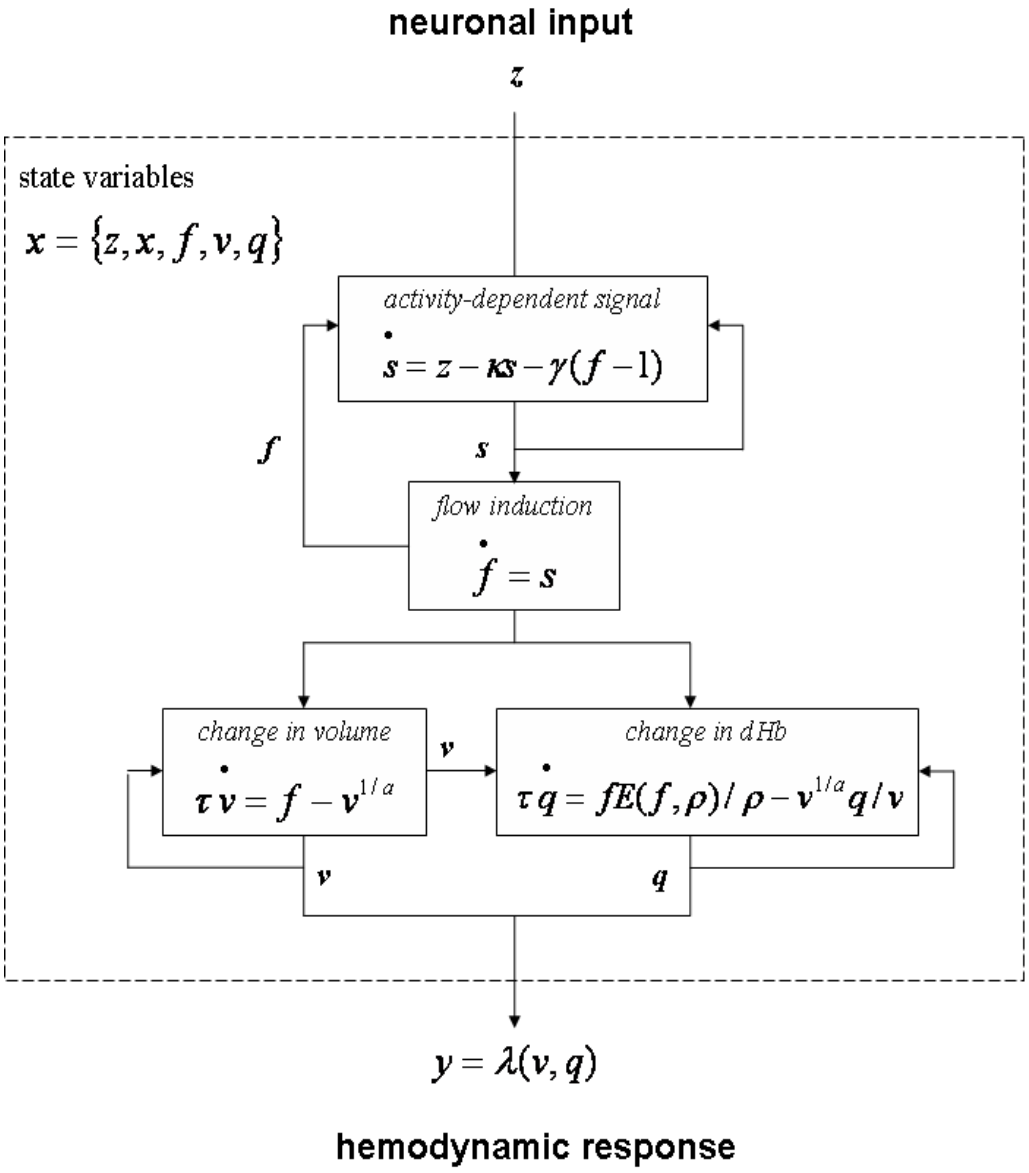


Figure 2.11: The architecture of hemodynamic model of a single region [12].

2.4.3 Granger Causality Mapping (GCM)

Granger causality mapping (GCM), a linear method developed for modelling time-resolved fMR time-series, investigates effective connectivity among activated brain areas using a vector autoregressive (VAR) model [14]. It is based on the concept of Granger causality, that is, causes always precede effects. Clive Granger first proposed the definition of causality for temporally structured data, which conceptually amounts to the following: if a time-series y causes (or has an influence on) x, then knowledge of y should help predict future values of x. Thus, causality (or influence) is framed in terms of predictability. Therefore, GCM aims at finding temporal precedence among time-series of brain regions to infer causal relationship, and computing the connectivity by evaluating interactions between a current voxel and a reference voxel. By modelling a time-series $x(t)$ as a vector autoregressive (VAR) process:

$$x(t) = -\sum_{i=1}^{p} A(i)x(t-i) + u(t) \tag{2.21}$$

where $u(t)$ is presumed to be multivariate white noise and p is the order of the auto-regression. The matrices $A(i)$ are the autoregressive (AR) coefficients as they regress $x(t)$ onto its own past.

Geweke [125] proposed a measure of linear influence $F_{x,y}$ between the time-series $x(t)$ and $y(t)$, which is regarded as an implementation of the concept of Granger causality in terms of VAR models. The influence measure $F_{x,y}$ is the sum of three components: the linear influence from x to y ($F_{x\to y}$), the linear influence from y to x ($F_{y\to x}$), and the instantaneous influence between x and y ($F_{x.y}$). The measure is defined using the residual cross-covariance matrices of the following three VAR models involving

time-series $x(t)$ and $y(t)$:

$$x(t) = -\sum_{i=1}^{p} A_x(i)x(t-i) + u(t) \qquad \text{var}(u(t)) = \Sigma_1 \tag{2.22}$$

$$y(t) = -\sum_{i=1}^{p} A_y(i)y(t-i) + v(t) \qquad \text{var}(v(t)) = T_1 \tag{2.23}$$

$$\text{and with} \qquad q(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} \tag{2.24}$$

$$q(t) = -\sum_{i=1}^{p} A_q(i)q(t-i) + w(t) \qquad \text{var}(w(t)) = Y = \begin{pmatrix} \Sigma_2 & C \\ C^T & T_2 \end{pmatrix} \tag{2.25}$$

where the residual cross-covariance matrices Σ_1, Σ_2, and Y quantify how well the current values of $x(t)$ and $y(t)$ can be predicted from their past values using linear AR models. The measures of total linear dependence between x and y, linear influence from x and y, linear influence from y to x, and instantaneous influence between x and y are defined, respectively:

$$F_{x,y} = \ln(|\Sigma_1| \cdot |T_1|/|Y|) \tag{2.26}$$

$$F_{x\to y} = \ln(|T_1|/|T_2|) \tag{2.27}$$

$$F_{y\to x} = \ln(|\Sigma_1|/|\Sigma_2|) \tag{2.28}$$

$$F_{x\cdot y} = \ln(|\Sigma_2| \cdot |T_2|/|Y|) \tag{2.29}$$

where "$|\Sigma|$"denotes the determinant of Σ. From these definitions, the following holds:

$$F_{x,y} = F_{x\to y} + F_{y\to x} + F_{x\cdot y} \tag{2.30}$$

Granger causality mapping is able to explore directed influence among brain regions in fMRI data without relying on *a priori* specification of an anatomical model that contains pre-selected regions and connections among them. Recently, a graphical ap-

proach linking the notions of graphical modelling and Granger causality is applied to describe the dynamic dependencies in neural systems [126]. Sparse MAR models are also proposed to deal with the problem of lacking degree of freedom in multivariate autoregressive (MAR) modelling [127]. However, the dynamics of the signal-fluctuations in GCM are driven by random error processes and the causes of the fluctuations are not explicitly modelled. Post hoc de-convolution analysis is required for GCM to ascertain stimulus-lock signal changes [128]. Moreover, in the context of more complicated bi-directional interactions, GCM tends to show only the dominant direction of influence.

2.5 Summary

This chapter provided the background of brain structures, function, and connectivity. Two earlier approaches for segmentation of brain sub-cortical structures based on deformable models and atlas-based models were discussed. The importance of spatial information and the use of brain templates in segmentation was emphasized. Previous work on the detection of brain activations from functional MR images were reviewed. The advantages and disadvantages of hypothesis-driven methods and data-driven methods were discussed. Lastly, three previous techniques for inference of effective brain connectivity namely, structural equation modelling (SEM), dynamic causal modelling (DCM), and Granger causal mapping (GCM), were presented and discussed.

Chapter 3

Modelling Structural Variability in MRI

Though the brain anatomy is roughly the same for all individuals, there are variations in the morphology of brain structures and tissues among subjects. Segmentation and quantification of subcortical structures of the human brain have important applications in clinical practice. In this chapter, we present a novel method to automatically segment subcortical structures of the human brain from magnetic resonance images. In order to deal with structural variability among individuals, a set of fuzzy templates of structures based on features such as intensity, spatial location, and relative spatial relationship are created automatically from a set of training images. The segmentation of a given image is performed by registering fuzzy templates and using tissue maps. The method does not require any specific expert guidance or manual interaction during the segmentation process. The chapter begins with our motivations followed by a description of the proposed methodology. The method is demonstrated by segmenting five structures: thalamus, putamen, caudate, hippocampus, and amygdala, from a set of real MR images. The performance of the present method is comparable or superior to earlier approaches.

3.1 Introduction

Magnetic resonance imaging (MRI) has become a significant imaging modality in clinical diagnosis and treatment planning. It is capable of providing detailed information of normal or diseased anatomy for medical research. Segmentation of subcortical structures from MR brain images is a critical task having numerous applications in quantitative morphometry, 3D volume visualization, mapping brain structures and functions, and clinical investigations including pathology, diagnosis, therapy, surgery planning and so on. Because of the lack of clear boundaries on MR images, manual segmentation of subcortical structures is time-consuming, subjective, and error prone. Inter- or intra-observer variability reduce the ability to detect subtle differences and to make comparisons in group studies. Moreover, the growing number of high resolution MR images have necessitated the use of computers to facilitate automatic segmentation [5, 4].

There are several issues that should be addressed by any automated subcortical structure segmentation technique. Inhomogeneous intensity within one tissue class or structure and overlapping intensity characteristics among different structures affect the effectiveness of intensity-based methods [129]. Although the spatial information has been incorporated to solve inherent ambiguities of class intensity distributions, because of the complexity and variability of the shape, size, and orientation, accurate identification of brain anatomical structures poses many challenges. Furthermore, low contrast, noise perturbation, partial volume effect, and magnetic field inhomogeneities of MR images should be carefully taken into account by any automated technique [4, 59].

A wide variety of methods have been proposed in the literature for segmentation of subcortical structures [5]. Deformable models deform a template based on extracted image features and have been extensively studied and widely used in medical image

segmentation with promising results [4, 67]. However, such methods rely on human experts for initialization and guidance. As prior knowledge from brain atlases is helpful to guide the segmentation process, elastic image registration techniques using brain atlas have been used to identify brain structures [56, 57]. However, imperfect registration largely influences the performance of segmentation. Bayesian approach has been applied to detect neuroanatomical structures [58] by the maximum a posteriori (MAP) estimation. A set of manually labelled training images are used to create probabilistic prior maps of structures by a linear registration to the atlas. Knowledge-based approaches such as fuzzy modelling [59] and information fusion [46] have also been adopted. Hybrid techniques were proposed to further increase the accuracy of detection, such as maximum a posteriori (MAP) estimation of structures with level set prior information [54], combining fuzzy clustering technique with deformable models for thalamus segmentation [61], integration of fuzzy spatial relations in deformable models [62], anatomically constrained region deformation based on prior knowledge of anatomical features [67], and fitting a group of deformable templates supervised by a series of rules derived from analyzing the template's dynamics and expert knowledge [130]. Shape descriptors such as moments of 3D coordinates were recently proposed to characterize the morphology of brain structures [131]. Other approaches include histogram analysis [63], neural network techniques [64, 65], knowledge-based approaches [68], and genetic algorithms [66]. Despite all these efforts, due to the limited efficacy of current automated techniques, clinicians still stick to manual segmentation of brain subcortical structures from MR images. There is a need to further enhance existing techniques or investigate into novel approaches that can automatically and accurately detect the subcortical structures of the brain.

We approach the problem of segmentation of subcortical structures by (1) incorporating spatial information of the structures as their intensities are not sufficient for reliable detection; (2) taking into consideration relative spatial relationships among

structures, which are more fixed and predictable though individual variability of the structures exists in terms of location, shape, size, and orientation; and (3) using a set of training images to obtain prior information of each structure of interest in the form of fuzzy templates as the information provided by a test image or a single rigid template is incapable of dealing with individual variabilities.

The expert knowledge such as "the putamen was defined as a gray matter prismatic-shaped structure, located approximately at 25 mm of the lateral ventricles, almost symmetric with respect to the interhemispheric plane and slightly posterior to the frontal horn of the lateral ventricles"may be useful for the segmentation of a certain structure [46]. Nevertheless, this kind of expert definition of brain subcortical structure may be vague and subjective; some parameters like "25mm"may not be suitable for a different group of subjects. Therefore, the present approach precludes expert definitions, instead, it automatically extracts the prior knowledge of structures from a set of manually segmented training images and encodes them into fuzzy templates. On the other hand, atlas-based approaches usually represent prior knowledge with a rigid template, which cannot take into account the variability of individual anatomy. As a result, we use fuzzy logic to model anatomical variability and postulate fuzzy templates as an appropriate way to represent prior knowledge of subcortical structures.

In order to handle the imprecision and the uncertainty inherent in brain structures and MR images, fuzzy templates of different subcortical structures in terms of features such as intensity, spatial location, relative distance, and relative direction, are first generated from a set of training images. The fuzzy templates are then used as the prior knowledge for segmentation by nonlinearly registering with the test image. Further, the information of tissues of the test image is used in conjunction with the fuzzy templates to form fuzzy maps of structures. A defuzzification step with α-cut on fuzzy maps is finally used to precisely localize and determine the boundary of sub-

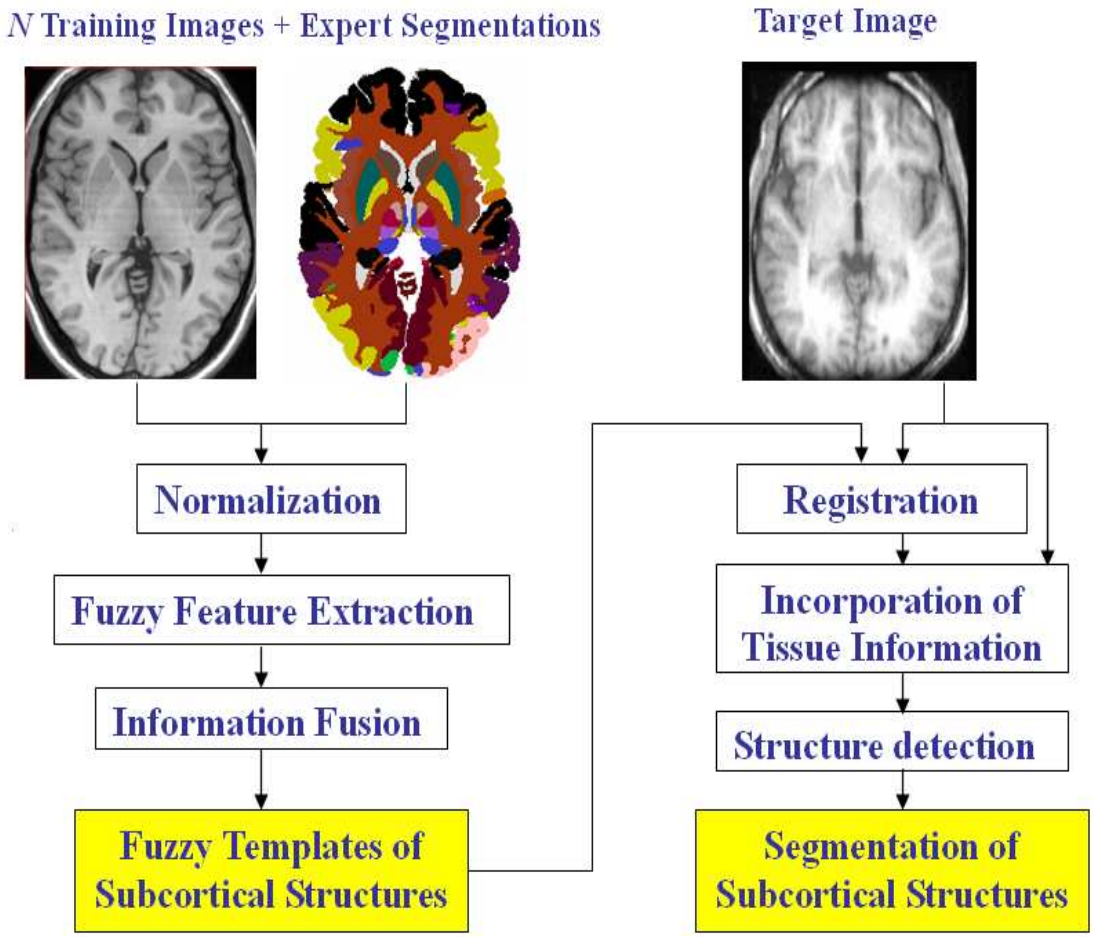


Figure 3.1: A schematic diagram illustrating the proposed fuzzy templates approach for segmentation of subcortical structures from MRI.

cortical structures. Segmentation of five subcortical structures: thalamus, caudate, putamen, hippocampus, and amygdala, on 17 images were performed to demonstrate the methodology; the results were compared with earlier approaches to illustrate its efficacy.

3.2 Method

The first step in the present method is to construct fuzzy feature templates of each structure using a set of training images normalized to a standard template. Fuzzy maps of each structure for a test image are then created by registering and fusing

the fuzzy templates and tissue maps. The final detection is made by a defuzzification step. A schematic diagram illustrating the method is shown in Fig. 3.1.

3.2.1 Construction of Fuzzy Templates

Suppose that there is a standard template $r = (r_i : i \in \Omega)$ of the human brain where r_i is the image intensity at voxel site i of template r and $\Omega \subset \mathbf{N}^3$ is the three-dimensional spatial domain of image voxels. A set of N real value 3D MR brain images $F = \{f^j : j = 1, 2, ...N\}$ along with their expert segmentations are available to learn the fuzzy templates of the structures. The j th training image is denoted by $f^j = (f_i^j : i \in \Omega)$ where f_i^j is the image intensity of voxel i in image f^j. Suppose that the expert segmentation identifies L number of structures indexed by the set $\Gamma = \{1, 2, ...L\}$ from each image and let the labelled image (expert segmentation) corresponding to image f^j be denoted by $s^j = (s_i^j = l : i \in \Omega, l \in \Gamma)$, in which voxel site i in image f^j belongs to the structure l.

Normalization

In order to construct fuzzy templates, all training images F are normalized to a standard template r resulting in a set of normalized images $\tilde{F} = \{\tilde{f}^j : j = 1, 2, ...N\}$. A 12-parameters affine transformation followed by nonlinear warping [85] is applied to standardize all brain scans to the same size and orientation under the same coordinate system. The optimal 12 parameters for registration can be computed by Gauss-Newton optimization algorithm as described in [70]. Nonlinear warping based on a linear combination of discrete cosine transform (DCT) basis functions corrects gross differences in head shapes among subjects. This optimization is extended to a maximum a posteriori (MAP) estimation regularized by the prior knowledge [82]. The same transformation obtained from normalizing training images f^j to the template r

is applied to the corresponding expert segmentation s^j to produce a set of normalized segmented images $\tilde{s}^j = (\tilde{s}_i^j = l : i \in \Omega, l \in \Gamma)$ for all $j = 1, 2, ...N$.

Fuzzy Feature Maps

For each pair of normalized segmentation $\tilde{s}^j$ and normalized image $\tilde{f}^j$, there are L groups of voxels that belong to different structures $l \in \Gamma$. Let us denote the set of voxels belonging to the structure l in the normalized image $\tilde{f}^j$ as $G_j^l = (i : \tilde{s}_i^j = l, i \in \Omega)$. For each structure $l \in \Gamma$ in each image $\tilde{f}^j \in \tilde{F}$, three fuzzy feature maps are created:

1. Intensity: We model the intensity characteristics of the structure l in the normalized training image $\tilde{f}^j$ as a fuzzy set, whose fuzzy membership function is estimated from the intensity histogram of G_j^l. For every voxel i in the normalized training image $\tilde{f}^j$, the fuzzy membership value A_{ij}^l of intensity to the structure l is calculated from the normalized training image $\tilde{f}^j$ and the segmentation $\tilde{s}^j$ as follows:

$$A_{ij}^l = \begin{cases} \frac{H^{int}(i,l,j) - H_{min}^{int}(l,j)}{2(H_{max}^{int}(l,j) - H_{min}^{int}(l,j))} + 0.5 & \text{if } \tilde{s}_i^j = l; \\ 0 & \text{otherwise,} \end{cases} \tag{3.1}$$

where $H^{int}(i,l,j)$ represents the number of voxels having the same intensity as voxel i in structure G_j^l of image $\tilde{f}^j$; $H_{min}^{int}(l,j) = \min_i H^{int}(i,l,j)$ and $H_{max}^{int}(l,j) = \max_i H^{int}(i,l,j)$ stand for the minimum and maximum number of voxels having the same intensity in G_j^l, respectively. Thus, for voxels belonging to G_j^l, the fuzzy membership of intensity is in the range $[0.5, 1]$ depending on its value in the histogram, otherwise the fuzzy membership value of zero is assigned. The above fuzzy membership function is designed to differentiate largely between those voxels belonging to a particular structure and other voxels, and to give higher fuzzy memberships to more prominent voxels and lower values to less prominent ones, based on the intensity distribution.

2. Spatial location: Suppose x_{ij}, y_{ij}, and z_{ij} stand for the spatial location of voxel i in the normalized image $\tilde{f}^j$ from left to right, from posterior to anterior, and from inferior to superior, respectively. The spatial location of structure l in $\tilde{f}^j$ is modelled with a three-dimensional fuzzy map based on histograms of x, y, and z values of all voxels belonging to the structure. Similar to the fuzzy membership function of intensity, the fuzzy membership function of spatial location in one direction D is defined as:

$$B(D)_{ij}^{l} = \begin{cases} \frac{H^D(i,l,j) - H^D_{min}(l,j)}{2(H^D_{max}(l,j) - H^D_{min}(l,j))} + 0.5 & \text{if } \tilde{s}_i^j = l; \\ 0 & \text{otherwise,} \end{cases} \tag{3.2}$$

where $D \in \{x, y, z\}$ stands for any of the three directions, $H^D(i, l, j)$ is the number of voxels in G_j^l having the same spatial location value in the direction D as voxel i in image $\tilde{f}^j$, $H^D_{min}(l, j) = \min_i H^D(i, l, j)$, and $H^D_{max}(l, j) = \max_i H^D(i, l, j)$. The combination of the three fuzzy memberships produces the spatial location fuzzy map of each structure l in each image $\tilde{f}^j$ as:

$$B_{ij}^{l} = \sqrt[3]{B(x)_{ij}^{l} B(y)_{ij}^{l} B(z)_{ij}^{l}} \tag{3.3}$$

3. Relative spatial relations: Since the topological relationships among subcortical structures in the human brain are predictable, relative spatial relations between structure G_j^l, $l \in \Gamma$ and other structures $G_j^{l'}$, $l'(\neq l) \in \Gamma$ in j th normalized training image can be stably expressed. Fuzzy maps of relative spatial relations based on relative distance $C(dis)_{ij}^{ll'}$ and relative direction $C(dir)_{ij}^{ll'}$ among structures are estimated from the distance and direction histograms, respectively. Let $\mathbf{v}_{ij}^{l'}$ denotes the vector starting from the centroid $o_j^{l'}$ of the structure l' to the voxel i of the image $\tilde{f}^j$. Then $d_{ij}^{l'} = |\mathbf{v}_{ij}^{l'}|$ stands for the distance between these two voxels in image $\tilde{f}^j$. The fuzzy

map of relative distance is given by:

$$C(dis)_{ij}^{ll'} = \begin{cases} \frac{H^{dis}(d_{ij}^{l'},l',l,j)-H_{min}^{dis}(l',l,j)}{2(H_{max}^{dis}(l',l,j)-H_{min}^{dis}(l',l,j))}+0.5 & \text{if } \tilde{s}_i^j = l; \\ 0 & \text{otherwise,} \end{cases} \tag{3.4}$$

where $H^{dis}(d_{ij}^{l'}, l', l, j)$ is number of voxels in G_j^l having same distance $d_{ij}^{l'}$ to the centroid $o_j^{l'}$ of structure l' in image $\tilde{f}^j$ from voxel i; $H_{min}^{dis}(l', l, j)$ and $H_{max}^{dis}(l', l, j)$ denote the minimum and maximum numbers of voxels in G_j^l at the same distance to the centroid $o_j^{l'}$ of structure l' in image $\tilde{f}^j$, respectively.

For relative direction modelling, a vector is formed from the centroid $o_j^{l'}$ of structure l' to voxel i in image $\tilde{f}^j$, denoted as $\mathbf{v}_{ij}^{l'}$. The direction of this vector is decided by the three angles made with xy-plane ($z = 0$), xz-plane ($y = 0$) and yz-plane ($x = 0$), respectively:

$$\alpha(xy)_{ij}^{ll'} = \begin{cases} arcsin(\frac{\mathbf{v}_{ij}^{l'} * (0\ 0\ 1)}{|\mathbf{v}_{ij}^{l'}|})\frac{180}{\pi} & \text{if } \tilde{s}_i^j = l, \\ 0 & \text{otherwise;} \end{cases} \tag{3.5}$$

$$\alpha(yz)_{ij}^{ll'} = \begin{cases} arcsin(\frac{\mathbf{v}_{ij}^{l'} * (1\ 0\ 0)}{|\mathbf{v}_{ij}^{l'}|})\frac{180}{\pi} & \text{if } \tilde{s}_i^j = l, \\ 0 & \text{otherwise;} \end{cases} \tag{3.6}$$

$$\alpha(xz)_{ij}^{ll'} = \begin{cases} arcsin(\frac{\mathbf{v}_{ij}^{l'} * (0\ 1\ 0)}{|\mathbf{v}_{ij}^{l'}|})\frac{180}{\pi} & \text{if } \tilde{s}_i^j = l, \\ 0 & \text{otherwise.} \end{cases} \tag{3.7}$$

where * indicate the inner product of the two vectors. The range of the above angles are $[-90, 90]$ and the fuzzy membership of relative direction of voxel i in image $\tilde{f}^j$ for all, $l, l' \in \Gamma$ and $i \in \Omega$ is given by

$$C(dir)_{ij}^{ll'} = \sqrt[3]{\prod_{\psi=\{xy,yz,xz\}} C(\psi)_{ij}^{ll'}} \tag{3.8}$$

where the form of the fuzzy membership function is same for three angles, $\psi \in$

$\{xy, yz, xz\}$ as below:

$$C(\psi)_{ij}^{ll'} = \begin{cases} \frac{H^{\psi}(\alpha(\psi)_{ij}^{ll'}, l, l', j) - H_{min}^{\psi}(l, l', j)}{2(H_{max}^{\psi}(l, l', j) - H_{min}^{\psi}(l, l', j))} + 0.5 & \text{if } \tilde{s}_i^j = l; \\ 0 & \text{otherwise,} \end{cases} \tag{3.9}$$

where $H^{\psi}(\alpha(\psi)_{ij}^{ll'}, l, l', j)$ is the number of voxels in G_j^l having the same angle $\alpha(\psi)_{ij}^{ll'}$ as voxel i to centroid $o_j^{l'}$ of structure l' with respect to direction ψ in image $\tilde{f}^j$; $H_{min}^{\psi}(l, l', j) = \min_i H^{\psi}(\alpha(\psi)_{ij}^{ll'}, l, l', j)$ and $H_{max}^{\psi}(l, l', j) = \max_i H^{\psi}(\alpha(\psi)_{ij}^{ll'}, l, l', j)$.

After obtaining fuzzy maps $C(dis)_{ij}^{ll'}$ (relative distance) and $C(dir)_{ij}^{ll'}$ (relative direction) for structure l with respect to every other structure l' in $\tilde{f}^j$, a fuzzy map describing the total relative spatial relations of structure l is obtained by fusion in two steps. For each voxel i in image $\tilde{f}^j$, we first combine relative distance and relative direction maps of structure l with respect to another structure l' followed by fusing all maps related to other reference structures l', $(l' \neq l)$, to produce a new fuzzy membership C_{ij}^l. We use T-norm fusion operator to obtain an overlap of information.

$$C_{ij}^{ll'} = \sqrt{C(dis)_{ij}^{ll'} * C(dir)_{ij}^{ll'}} \tag{3.10}$$

$$C_{ij}^{l} = \sqrt[L-1]{\prod_{l'=1, l' \neq l}^{L} C_{ij}^{ll'}} \tag{3.11}$$

In this way, every training image has three fuzzy maps for each structure $l \in \Gamma$ based on intensity, spatial location, and relative spatial relations: A_{ij}^l, B_{ij}^l, and C_{ij}^l where $i \in \Omega$ and $j = 1, 2, ..N$. Appendix A explains the detailed derivation of these fuzzy maps. Note that the final fuzzy membership of each voxel in every fuzzy map should be in the range $[0.5, 1]$ for true belongingness. Thus, we can differentiate the true voxels belonging to each structure from the others in one image more accurately. The resulting maps are better prepared for the next step of fusion to create the fuzzy

templates of the structures.

Information Fusion

Three different features - intensity (A), spatial location (B), and relative spatial relations (C) - are modelled by fuzzy feature maps of each training image, which are fused to obtain a total fuzzy template of the structure of interests.

A mean fusion operator is used to produce average fuzzy memberships of structure l for each voxel $i \in \Omega$, across all training images:

$$W_i^l = \frac{1}{N} \sum_{j=1}^{N} W_{ij}^l \tag{3.12}$$

where $W = \{A, B, C\}$ represents the set of features: intensity, spatial location, and relative spatial relations, respectively. Thereby, for each structure of interests $l \in \Gamma$, three fuzzy feature templates $A^l = (A_i^l : \ i \in \Omega)$, $B^l = (B_i^l : \ i \in \Omega)$, and $C^l = (C_i^l : i \in \Omega)$ were constructed. Furthermore, information fusion is applied to form the fuzzy template $T^l = (T_i^l : \ i \in \Omega)$ of structure l to include all three fuzzy features:

$$T_i^l = ((A_i^l * B_i^l)^{0.5} * C_i^l)^{0.5} \tag{3.13}$$

The reason for combining intensity and spatial locations first is to take into consideration overlapped information of intensity and location of all voxels belonging to a particular structure l. Incorporation of feature template of relative spatial relations, which is based on spatial locations of voxels both in structure l and other structures, further refines the information for accurate identification of structure l.

3.2.2 Segmentation

When a test image f is given, a three-steps scheme is employed to segment different structures $l \in \Gamma$ based on the fuzzy templates A^l, B^l, C^l, and T^l constructed using the training images.

Registration

In order to make use of fuzzy templates created during training, fuzzy membership information must be made available to each voxel in the test image f. Registration of fuzzy templates to the test image is performed by the same method as in the training stage but in the reverse direction. As a result, all voxels in the test image f have prior fuzzy memberships to structure $l \in \Gamma$ based on the registered fuzzy features. Let us denoted the fuzzy maps of the test image by $\widetilde{A}_i^l$ for intensity, $\widetilde{B}_i^l$ for spatial location, $\widetilde{C}_i^l$ for relative spatial relation, and $\widetilde{T}_i^l$ for total information.

Incorporation of Tissue Information

The brain matter can be broadly classified into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). Different subcortical structures have different tissue properties, which information is proved to be useful for segmentation [132]. Therefore, fuzzy maps of three tissue classes $k \in \{$ GM, WM, CSF $\}$ are constructed to grasp more unique properties of different structures in the test image f such as shape, size, and orientation, which may not be satisfactorily captured by the registered fuzzy templates.

There are many approaches for tissue segmentation and we used an approach proposed by Ashburner et al. [84], which is an extension of a clustering algorithm based on the maximum likelihood mixture model. It incorporates a prior spatial probability map

of belonging to each tissue type and uses a Bayesian model to correct image intensity non-uniformity. The process starts by estimating probabilities of each tissue class based on *a priori* map and then iteratively computes cluster parameters, belonging probabilities, and sensitivity fields from intensity corrected images until a convergence criterion is achieved [84, 73]. A mixture of Gaussian functions is used to model the probability distribution of three tissue classes, which can be seen as fuzzy membership functions with Gaussian shapes. Therefore, the final probabilities of belonging for each voxel to a particular tissue class produced by this segmentation algorithm can be treated as fuzzy membership values.

Until now, two kinds of information are available for segmentation purpose: (1) the prior knowledge of intensity, spatial location, and relative spatial relations of one structure l, rendered by the registered fuzzy templates $\widetilde{A}^l$, $\widetilde{B}^l$, $\widetilde{C}^l$ and $\widetilde{T}^l$; and (2) the tissue segmentation of the test image itself denoted by $P = (P_i^k : \ i \in \Omega, k = \{$ GM, WM, CSF $\}$, where P_i^k stands for the probability of belonging to tissue class k and its value is in the range $[0, 1]$. The total fuzzy template $\widetilde{T}^l$ consisting of all information in three feature templates is combined with the tissue information map P to produce a more accurate fuzzy membership map of structure l in the test image f. T-norm operator is applied for fusion as follows:

$$Z_i^l = \sqrt{\widetilde{T}_i^l P_i^k} \text{ if structure } l \text{ contains tissue type } k. \tag{3.14}$$

where Z_i^l is the fuzzy membership value giving the degree of belongingness of voxel i to the structure l.

Detection of Structures

Decision or defuzzyficaton is required to decide on the belongingness of each voxel i to structure l in the test image f. For each voxel i, we find the structure l^* with

the maximum fuzzy membership among all structures of interests $l \in \Gamma$, and assign zero fuzzy membership value to all other structures ($l \neq l^*$) for this voxel, to avoid overlapping that is,

$$\text{if } l^* = \arg\max_{l \in \Gamma} Z_i^l \text{ then let } Z_i^l = 0 \text{ for } l \neq l^* \tag{3.15}$$

We apply a simple α-cut thresholding technique to this final fuzzy map Z. Voxel i in image f belongs to structure class l if

$$Z_i^l \geq \xi_l \tag{3.16}$$

where threshold ξ_l is automatically learned from all training images of structure l. The resulting group of voxels belonging to structure l are the most confident group satisfying all the information above. In other words, structure l has been identified from test image f based on intensity, spatial location, relative spatial relations, and tissue information.

Determination of the Threshold

This section describes the procedure used for determining the threshold ξ_l in Eq. 3.16 for detecting the structure l from a given test image f. Suppose a fuzzy template T^l was created from the set F of N training images in order to identify structure l. During training, an optimal threshold ξ_l^j for detecting the structure l in each training image f_j is learned by using the combined fuzzy map $Z^{lj} = (Z_i^{lj} : i \in \Omega, j = 1, 2, ...N)$. The optimal threshold is then chosen in such a manner that the target structure is detected from the training image to have the highest mean of the volume difference and overlap indices computed against a known manual segmentation as the ground truth. Each training image produces one optimal threshold, resulting in a set of

thresholds $\{\xi_l^j : l \in \Gamma, j = 1, 2, ...N\}$. The confidence value of each threshold ξ_l^j is noted by weights $w_l^j = ({I_1}_l^j + {I_2}_l^j)/2$. As I_1 and I_2 are both in the range $[0, 1]$, the confidence weights $\{w_l^j : l \in \Gamma, j = 1, 2, \ldots N\}$ are in the same range as well. The threshold ξ_l for detecting the structure l in a given test image f is determined by taking a weighted average of the set of thresholds, ξ_l^j, obtained from the training images, with their corresponding confidence weights, w_l^j.

3.3 Experiments and Results

This section first introduces the set of MR brain images used in segmentation experiments and three quantitative indices used for evaluating performance of the segmentation algorithm. The experiments and results of comparison with other approaches are then given.

3.3.1 Data

MR brain images and their manual segmentations for our experiments were obtained from the Center for Morphometric Analysis at Massachusetts General Hospital [133]. There are 18 images in this data set; one image was identified to have relatively large ventricle system and hence removed from our analysis. Each of 17 T1-weighted volumetric images is composed of 128 coronal slices of 1.5mm thickness with 256×256 matrix per slice; voxel size is $0.94 \times 0.94 \times 1.5\text{mm}^3$. The images have been positionally normalized into the Talairach coordinates (rotation only) and been corrected for bias fields.

The standard reference template r for normalization of training image is the anatomical MRI brain template of a single subject provided by International Consortium for Brain Mapping [29]. It has 362 coronal slices of 0.5mm thickness with 304x309 matrix

per slice and a voxel size of 0.5x0.5x0.5mm^3. Since our method is voxel-based, this standard template is used as a reference to create fuzzy feature maps and to combine the different features together to create a single fuzzy template for each structure.

3.3.2 Quantitative Indices for Validation

Three quantitative measures were used to evaluate the performance of the proposed segmentation algorithm with reference to the expert segmentations [55, 56, 46]. Suppose that the true group of voxels for structure l is denoted by G_t^l, the group of voxels for structure l produced by a segmentation algorithm is G_c^l, and the $V(G)$ returns the volume or the number of voxels in the group of voxels G.

1. Percent volume difference is computed from the relative error in volume estimation when the true volume is given by the expert:

$$I_1 = 1 - \frac{|V(G_t^l) - V(G_c^l)|}{V(G_t^l)} \tag{3.17}$$

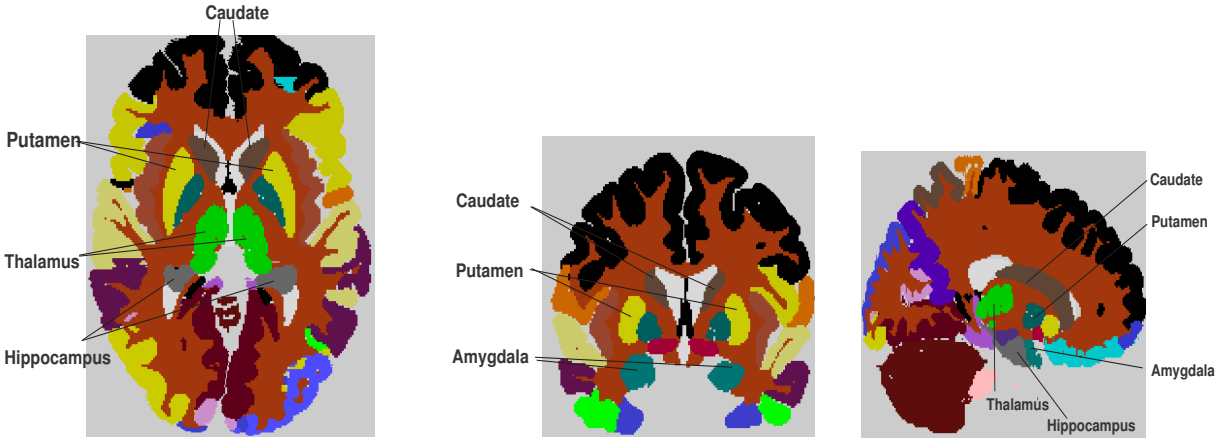
2. Percent overlap or spatial accuracy assesses the relative overlapping of the computed structure G_c^l and the true structure G_t^l as:

$$I_2 = \frac{V(G_t^l \cap G_c^l)}{V(G_t^l)} \tag{3.18}$$

3. Mean distance measures the absolute distance (in millimeters) of the computed structure to the true structure, defined as:

$$I_3 = \frac{\sum_{c \in G_c^l} \min_{t \in G_t^l} ||c - t||}{V(G_c^l)} \tag{3.19}$$

where c represents a voxel in the detected structure G_c^l, and t represents a voxel in



(a) Axial view (b) Coronal view (c) Sagittal view

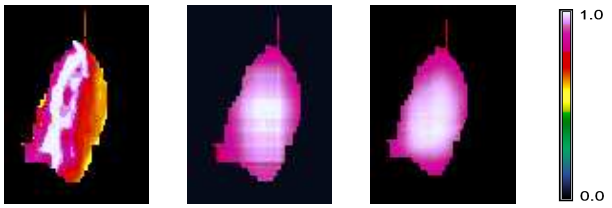
Figure 3.2: Illustration of location and shape of five subcortical structures of interests.

the true structure G_t^l; $||c-t||$ stands for the distance between two voxels.

Perfect matching between the computed volume and true volume is achieved when the first two performance measures equal to one and the mean distance measured equals to zero. Note that some published results have used the average of the predicted volume and manually labelled volume [58, 59] or the union of the two [61] as the normalization factor, as opposed to the manually labelled volume used here.

3.3.3 Creating Fuzzy Templates

In this section, the creation of fuzzy templates is demonstrated using a set of training images. For illustration purpose, the location and shape of all five structures in left and right brain is shown on axial, sagittal, and coronal slices in Fig. 3.2. The whole process is in 3D, but some 2D slices are shown for illustration purposes. Different fuzzy feature maps of structure l based on a single training image f_j are first calculated. For example, Fig. 3.3 presents (a) the intensity fuzzy map A_j^l, (b) the spatial location fuzzy map B_j^l, and (c) the relative spatial relations fuzzy map C_j^l of structure l (left thalamus) for a representative training image f_j.

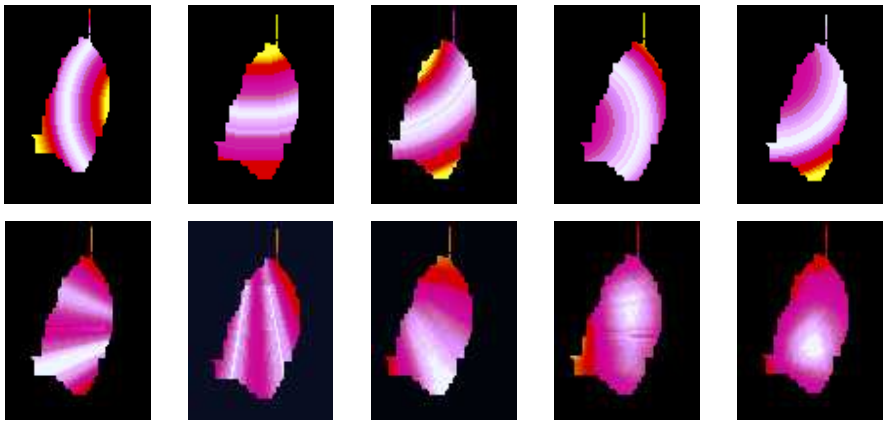


(a) (b) (c)

Figure 3.3: Fuzzy feature maps of left thalamus on an axial slice ($z = 155$) in one representative training image: (a) intensity fuzzy map; (b) spatial location fuzzy map; and (c) relative spatial relations fuzzy map. The color bar represents color coding of fuzzy membership, in which the upper end represents one and the bottom end represents zero.

For creating relative spatial relations fuzzy map C_j^l as shown in Fig. 3.3(c), five reference structures are chosen for each target structure, which includes four other structures on the same side of the brain and the same target structure in the opposite side of the brain. For instance, to consider the relative spatial relations of left thalamus, the corresponding reference structures are putamen, caudate, hippocampus, and amygdala in the left brain and thalamus in the right brain. Relative spatial relations are computed in terms of relative direction and relative distance. Fig. 3.4 shows relative distance (first row) and relative direction (second row) fuzzy maps with respect to five reference structures for structure l (left thalamus) in one representative training image. These fuzzy maps are combined together by Eq. 3.8 - 3.11 to form one fuzzy map representing relative spatial relations of left thalamus C_j^l for the training image f_j. Other nine structures follow exactly the same procedure to produce fuzzy maps.

After that, the fuzzy feature maps of all the training images are fused together to create a total fuzzy template of subcortical structures by Eq. 3.12 and 3.13. Fig. 3.5 presents three fuzzy feature templates based on intensity A^l, spatial location B^l, and spatial relationships C^l, respectively, resulting from fusion of all training images. Fig. 3.6 presents the total fuzzy templates T^l, $l \in \Gamma$, for ten structures including thalamus, putamen, caudate, hippocampus, and amygdala (in both left and right).



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

Figure 3.4: Fuzzy maps of relative spatial relations of left thalamus on an axial slice ($z = 155$) in one representing training image. First row show relative distance fuzzy maps; second row give relative direction fuzzy maps. They are calculated relative to (a) right thalamus; (b) left caudate; (c) left putamen; (d) left hippocampus; and (e) left amygdala. The same color bar measure was used as in Fig. 3.3.

3.3.4 Segmenting Subcortical Structures

The fuzzy membership of a specific voxel represents its degree of belongingness to the particular structure. This kind of prior knowledge is used in the segmentation step as described in section 3.2.2, and is fused with gray matter tissue fuzzy map. All five structures of interests are made out of brain gray matter tissue. The final combined fuzzy map is able to delineate the target structure by drawing fuzzy boundaries. The decision is made to accurately identify the structure by first computing the structure which a voxel has the maximum belongingness. And thereafter retain voxels having a membership above a certain threshold ξ_l^{test} in the combined fuzzy map of the structure l of the test image. The thresholding is performed on the three-dimensional image volume. As described in section 3.2.2, the threshold for identifying structure l from the test image was automatically learned during the detection of structure l from all the training images. Post-processing to remove unconnected voxels is applied at the final decision.

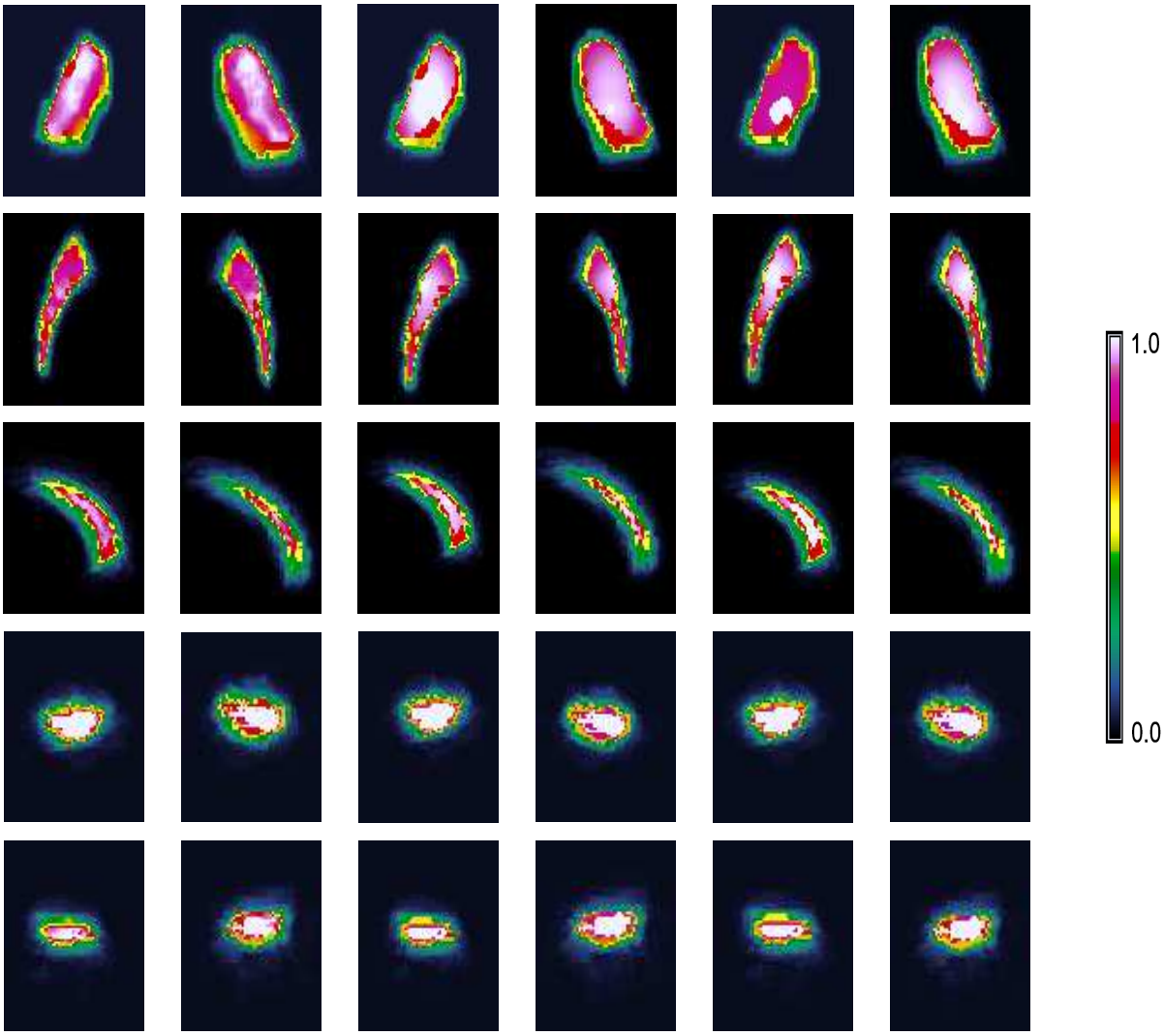


Figure 3.5: Total fuzzy feature templates based on intensity (a)(b), spatial location (c)(d), and relative spatial relationships (e)(f) for ten structures of interests (row 1 to 5): thalamus, putamen, caudate, hippocampus, and amygdala (left brain: (a), (c), and (e); right brain: (b), (d) and (f)). Thalamus (axial slice $z = 155$); putamen (axial slice $z = 136$); caudate (sagittal slice $x = 177$ for left and $x = 122$ for right); hippocampus (coronal slice $y = 184$); amygdala (axial slice $z = 103$). The color bar represents the range of fuzzy memberships, in which the top end represents one and the bottom represents zero.

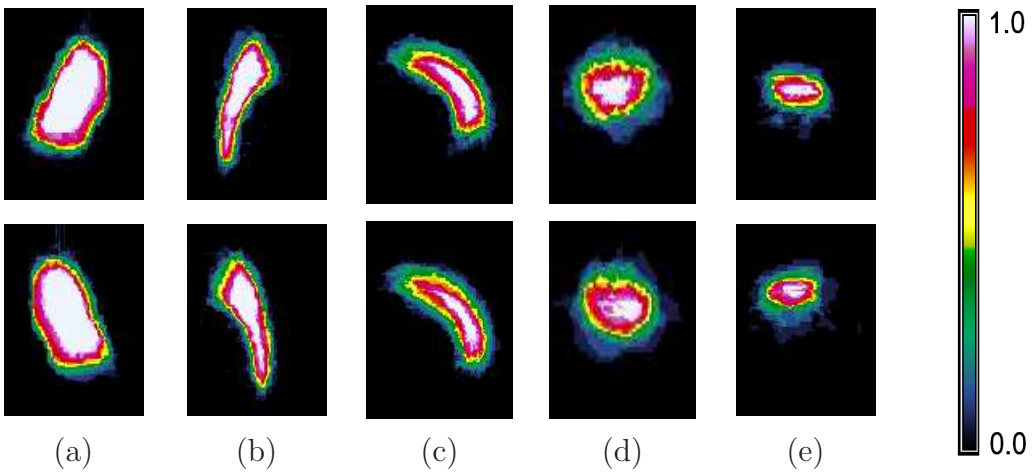


Figure 3.6: Total fuzzy templates for ten structures of interests: (a) thalamus, (b) putamen, (c) caudate, (d) hippocampus, and (e) amygdala. First and second row show structures in the left and right brain, respectively. Thalamus (axial slice $z = 155$); putamen (axial slice $z = 136$); caudate (sagittal slice $x = 177$ for left and $x = 122$ for right); hippocampus (coronal slice $y = 184$); amygdala (axial slice $z = 103$). The color bar represents the range of fuzzy memberships, in which the top end represents one and the bottom represents zero.

3.3.5 Cross-Validation

In order to validate our approach, leave-one-out (LOO) cross-validation method was used on a data set of 17 images described in section 3.3.1. In cross-validation cycle, fuzzy templates were obtained by training 16 images and then tested on the left-out image. The results are then evaluated by comparing to the manual segmentation in terms of the three performance indices described in section 3.3.2. Table 3.1 gives the best, worst, and average segmentation results for all 17 images in terms of the three performance indices of ten structures - thalamus, putamen, caudate, hippocampus, and amygdala (both left and right) detected by the present approach. Fig. 3.7 shows the best and worst segmented structures on one representative brain slice. The true volume and predicted volume are shown together with the overlapping part. Note that the experts did the manual segmentation in coronal slices, leading to smooth edges in coronal slices but chaotic edges in axial and sagittal slices. Because only selected 2D slice is shown, the connected part of the structure in 3D may be seem as

		Left			Right		
Structure	Index	Best	Worst	Mean	Best	Worst	Mean
Caudate	$I1$	0.99	0.71	0.91 ± 0.07	0.88	0.85	0.90 ± 0.07
	$I2$	0.80	0.82	0.81 ± 0.06	0.89	0.64	0.80 ± 0.08
	$I3$	0.25	0.50	0.31 ± 0.11	0.26	0.55	0.32 ± 0.11
Putamen	$I1$	0.98	0.72	0.91 ± 0.08	0.99	0.87	0.91 ± 0.08
	$I2$	0.90	0.87	0.84 ± 0.06	0.84	0.68	0.81 ± 0.06
	$I3$	0.13	0.57	0.25 ± 0.12	0.20	0.35	0.33 ± 0.12
Thalamus	$I1$	0.95	0.88	0.95 ± 0.03	0.98	0.99	0.95 ± 0.03
	$I2$	0.89	0.75	0.83 ± 0.05	0.88	0.69	0.84 ± 0.06
	$I3$	0.23	0.22	0.32 ± 0.17	0.19	0.82	0.32 ± 0.19
Hippocampus	$I1$	0.95	0.85	0.89 ± 0.09	0.95	0.89	0.88 ± 0.07
	$I2$	0.86	0.60	0.72 ± 0.11	0.77	0.43	0.70 ± 0.11
	$I3$	0.23	1.17	0.55 ± 0.27	0.37	1.21	0.56 ± 0.24
Amygdala	$I1$	0.99	0.79	0.81 ± 0.17	0.99	0.76	0.78 ± 0.20
	$I2$	0.72	0.36	0.66 ± 0.15	0.73	0.41	0.64 ± 0.15
	$I3$	0.23	1.08	0.63 ± 0.25	0.39	1.03	0.70 ± 0.22

Table 3.1: Quantitative evaluation of segmentation of subcortical structures: best-case, worst-case, and average over 17 subjects. I_1 - percent volume difference, I_2 - percent volume overlap, and I_3 - mean distance for segmentation (mm); the best and worst cases were decided by the maximum and minimum values of the sum $I_1 + I_2$, respectively. The columns start with left and right in the first row denotes structures in the left and right brain, respectively.

unconnected.

A tradeoff between the indices I_1 and I_2 is seen. As the predicted volume is increasing, the percent volume difference index I_1 and percent volume overlap index I_2 keep increasing; but, when the predicted volume is larger than the true volume, I_1 starts to decrease. Different structures have different intensity, shape, size, and orientation variance. Therefore, the threshold learned from the training stage may not be the optimal one for the test image. Good segmentation was achieved for thalamus, putamen, and caudate with very low mean distance error, which is less than 0.5mm on average. This is an indication that the volume estimation is very consistent with respect to the expert quantification, and a good match between the expert segmentation and the results of our segmentation. Hippocampus and amygdala are rather small subcortical structures with significant individual variations, so the segmentation accuracy still

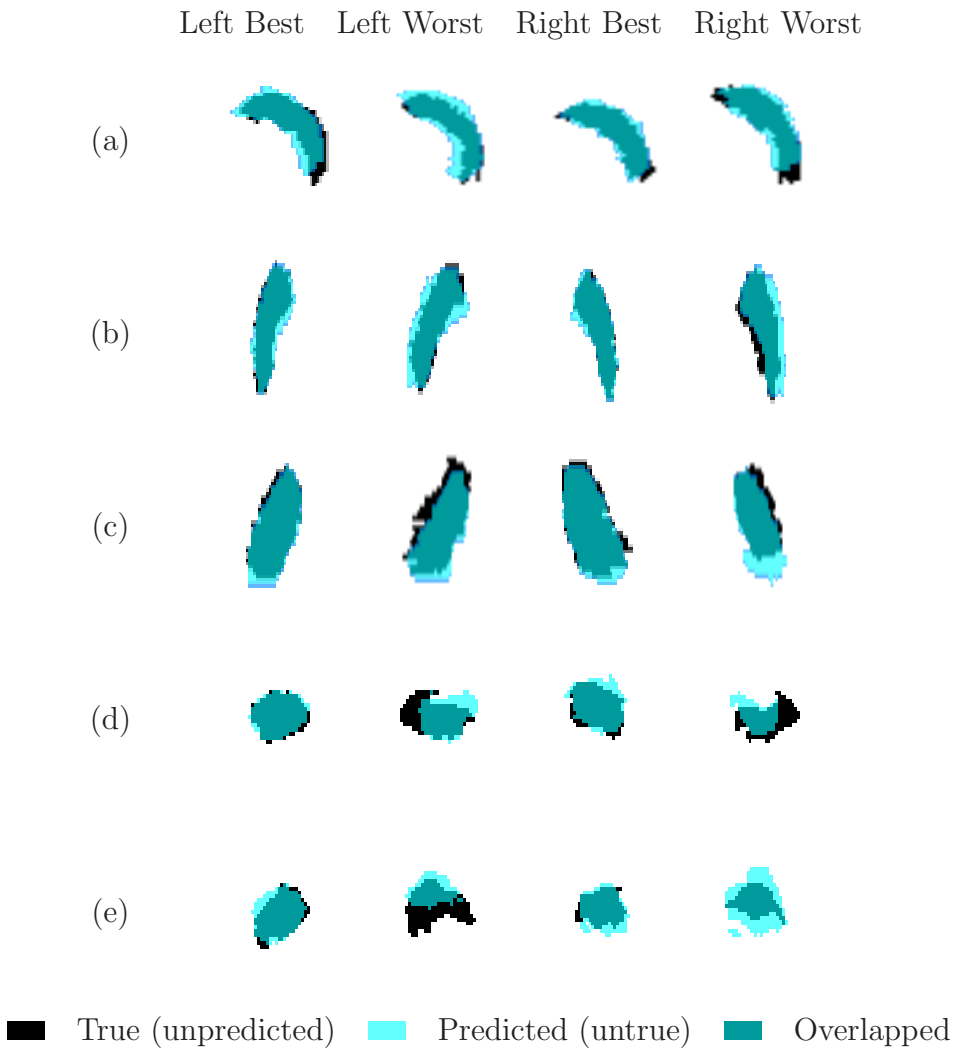


Figure 3.7: Best and worst segmentation results with reference to manual segmentation of ten subcortical structures on selected slices: (a) caudate (sagittal), (b) putamen (axial), (c) thalamus (axial), (d) hippocampus (coronal), and (e) amygdala (coronal). Columns (i) and (ii) show best and worst results for the structures in the left brain, while (iii) and (iv) columns show best and worst results for the structures in the right brain. The overlapped volume indicates where both the true volume and the predicted volume overlapped.

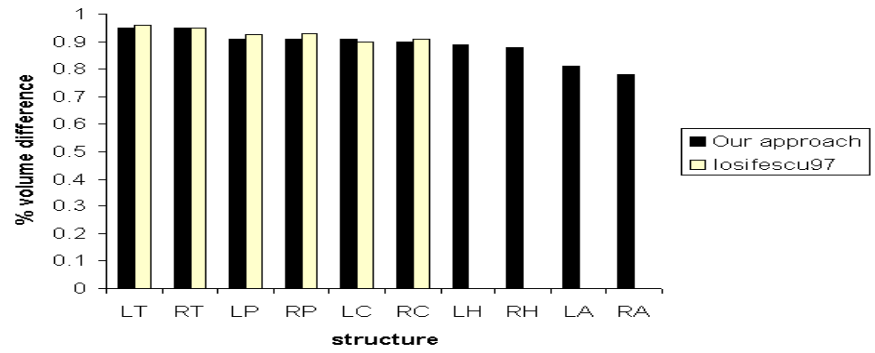
need improvement although the mean distance index were less than 1mm in most cases.

Fig. 3.8 compares the present segmentation results in terms of average accuracy for all ten structures with the approach presented in [55] using same indices (I_1, I_2 and I_3). Although the images used are different, we can still make some comparisons to see the potential of the present method. The elastic registration algorithm uses a MRI brain atlas based on a single subject [134], but our method creates fuzzy templates representing an ensemble of training images. We have comparable performance for thalamus and better accuracy for putamen and caudate. In particular, the volume overlap index is much higher, so the significance of the use of the fuzzy templates is justified. Furthermore, our approach can reasonably detect structures like amygdala and hippocampus. Our results for hippocampus is comparable to the segmentation results of individual images shown in [56] where no average results are given. In addition, the simulation was done on a Pentium IV 2.4GHz machine having 512MB memory and running MATLAB (version 7.0) in Windows XP environment. The execution time for the construction of fuzzy templates of all ten subcortical structures from 16 training images was approximately 75min, in which normalization of one image takes 3min. The segmentation of ten subcortical structures from one target MR image takes 11min. The running time is quite comparable with the method of [55], in which the entire process of automatic segmentation requires less than 90min of computer time on a SPARC workstation including 15min parallel segmentation. Note that the proposed approach is more applicable for clinical investigations because after fuzzy templates of subcortical structures have been constructed from the existing database of segmented images, the time to segment a new MR image is almost real-time.

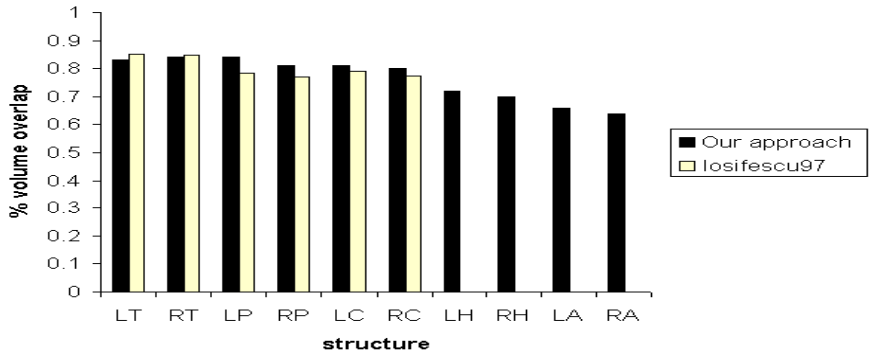
The method of [46], though based on a similar framework using fuzzy maps for the segmentation as ours, treats each structure individually by doing different fusion steps

with reference to expert definitions. Similarly, in [132], same framework is applied to segment subthalamic nucleus (ST). However, they first segmented the third ventricle (V3) and the two read nucleus (RN), and following identified subthalamic nucleus (ST) based on these neighboring structures by using topological information provided by expert and tissue information from the image. This procedure largely depends on whether the two neighboring structures are accurately identified. Moreover, some parameters in their expert definitions such as ˜25 mm˝between two structures are quite vague and cannot be generalized for other group of subjects. In short, we took a holistic approach for the detection of subcortical structures, in which all interested structures are detected simultaneously. Manual interactions or guidance were avoided during the segmentation process except the use of a set of pre-segmented images at the training stage. However, an individualistic approach, such as [46], may turn out to be more accurate for volume overlap measure for one group of subjects for a particular structure, but may not be accurate, in general, for other groups. Therefore, an objective comparison of our method with their approach is inappropriate.

After the current approach has been proposed, several different techniques has also been developed for sub-cortical structure segmentation. An intensity normalization procedure was proposed to improve the segmentation in atlas-based approach, which achieved comparable performance as our approach [57]. Chupin et al. also achieve comparable or higher performance for hippocampus and amygdala and their subparts using anatomically constrained deformable models [67]. Fuzzy spatial relations has also been proposed to incorporate into deformable models to segment caudate nucleus with better performance [62]. Due to different image and scanner properties of the used datasets, it is rather difficult to compare the methods objectively. However, all these approaches focus on one or few sub-cortical structures, and ours is a general method applicable to a wide range of sub-cortical structures.



(a) Percent volume differences of various structures



(b) Percent volume overlaps of various structures

Figure 3.8: Performance comparison of the present approach and the approach of [55]. Key: LT - left thalamus, RT - right thalamus, LP - left putamen, RP - right putamen, LC - left caudate, RC - right caudate, LH - left hippocampus, RH - right hippocampus, LA - left amygdala, RA -right amygdala.

3.3.6 Registration Artifacts

The registration step plays an important role in both training and testing. In order to evaluate the effect of the nonlinear registration algorithm on the performance of our approach, an experiment was designed involving the following two steps: (1) register both the test image f and its expert segmentation s to a standard template r, say by transform M', to obtain the normalized segmented images $\tilde{s} = M'(s)$; (2) register the same standard template r back to the test image by another transformation: $f = M(r)$, and apply it to $\tilde{s}$ to get $\tilde{\tilde{s}} = M(\tilde{s})$.

If the nonlinear registration algorithm is able to perfectly match the target image f and the template r, different anatomical parts of the brain should be aligned accordingly although sampling errors still exist, in other words, $\tilde{\tilde{s}}$ approximates s nicely for each structure l. However, usually this may not be the case because the presence of individual variability and registration artifacts. Therefore, for each structure l of interest, by comparing the expert segmentation s^l with the processed segmentation $\tilde{\tilde{s}}^l$, registration accuracy can be evaluated in terms of volume difference, volume overlap, and mean distance error as shown in Table 3.2. The registration error can be seen in terms of low volume difference, low volume overlap, and high mean distance error. Note that this registration error includes twice of the normalization step. The above process emulates the normalization procedure involved in the training and testing stage, so the errors and the information lost in the registration step influences the performance of our segmentation approach.

A simple mapping from a crisp segmented template to the target image is not sufficient for segmentation of subcortical structures because of registration artifacts. The fuzzy templates in our approach encode the necessary anatomical information in a 3D fuzzy membership distribution; hence, it is capable of adapting to the target image and identifying the structure of interest better than crisp template after registration.

Structure	I_1	I_2	I_3(mm)	$I_3^{r/s}$(%)
Caudate (L)	0.64 ± 0.07	0.53 ± 0.06	0.22 ± 0.19	70.9
Caudate (R)	0.65 ± 0.08	0.54 ± 0.07	0.23 ± 0.22	71.8
Putamen (L)	0.75 ± 0.08	0.63 ± 0.07	0.21 ± 0.11	84.0
Putamen (R)	0.74 ± 0.07	0.61 ± 0.06	0.25 ± 0.13	75.7
Thalamus (L)	0.76 ± 0.08	0.69 ± 0.06	0.14 ± 0.11	43.8
Thalamus (R)	0.75 ± 0.08	0.68 ± 0.05	0.13 ± 0.12	40.6
Hippocampus (L)	0.63 ± 0.08	0.51 ± 0.09	0.25 ± 0.13	45.4
Hippocampus (R)	0.62 ± 0.07	0.49 ± 0.08	0.30 ± 0.16	53.5
Amygdala (L)	0.66 ± 0.09	0.47 ± 0.09	0.40 ± 0.18	63.4
Amygdala (R)	0.68 ± 0.08	0.46 ± 0.09	0.49 ± 0.21	70.0

Table 3.2: Quantitative evaluation of the effect of registration on subcortical structures averaged over 17 subjects: I_1 - percent volume difference, I_2 - percent volume overlap, I_3 - mean distance in terms of millimeter, and $I_3^{r/s}$ - percentage of registration mean distance error divided by segmentation mean distance error. The bracket (L) and (R) denotes structures in the left and right brain, respectively.

Despite this, the registration still contributes to a significant amount of segmentation error as seen in the fourth column $I_3^{r/s}$ of Table 3.2. The highest ratios occurred in detecting putamen, which are 84.0% (L) and 75.7% (R); the lowest occurred for thalamus, which are 43.8% (L) and 40.6% (R). The higher ratios reflect that the accuracy of identification of a particular structure relies heavily on the registration procedure. Despite the heavy dependence on the registration procedure, the overall accuracies of the detection of subcortical structures from our algorithm were comparable or superior to the earlier approaches. However, more intelligent registration algorithms creating accurate fuzzy templates of these structures and enhancing the fuzzy templates' ability to adopt to a particular image will certainly improve the accuracy of our approach.

Furthermore, because of the nonlinear transformations used in the registration procedure, some voxels that belong to one structure in the original manual segmentation are seen as outliers on the final registered version of this structure. A connected component technique as a post-processing step to registration, was used in our approach to remove such unconnected voxels in each structure.

3.4 Discussion

A novel method for the segmentation of brain subcortical structures was presented. Fuzzy templates for different structures were constructed based on the information extracted from a set of training images. Important features such as intensity, spatial location, and relative spatial relations were considered and fused together. Fuzzy templates were registered to a given test image and further fused with its tissue information to produce fuzzy memberships of different structures. The final fuzzy maps successfully segmented thalamus, putamen, caudate, hippocampus, and amygdala in 17 T1-weighted MR images. Our method showed at least comparable or even better results compared to the previous approaches in the experiments, as measured by three quantitative indices: percent volume difference, percent volume overlap, and mean distance.

The proposed approach is based on a framework involving intensity, spatial location, relative spatial relations, and tissue features, and requires no specific expert definitions of structures or interventions during the segmentation process. The prior information of the structures was automatically extracted from the training images. The fuzziness of the maps captures the individual variability of each structure among the group of test images with the aid of fuzzy templates. Our approach is easily generalizable to other structures or to other image data, for handling the variability of structures and objects.

There are some limitations in the proposed method. The segmentation performance of one test image depends on the prior knowledge of the fuzzy template derived from a set of training images. Intensity, spatial location, and relative spatial relations vary among different group of subjects, and may change during development. For instance, intensity may change with age-dependent iron accumulation or myelination, spatial localization may vary with gender or other variables, and relative spatial relations

may change during development. Hence, it is necessary to consider the population under study and select a set of training images, that is a good representation of the population. Moreover, according to the experiments performed, registration step plays an important role both in training and testing stages. Any improvements of the registration could enhance the ability of fuzzy templates to adapt to a particular image and thus improve the accuracy of our technique.

We demonstrated the segmentation of only five subcortical structures (left and right). For the calculation of the fuzzy template of relative spatial relations of one structure, we chose to assign four other structures on the same side and the same structure in opposite side of the brain as the reference structures. However, certain structures may not have reference structures in all directions, and this affects the accuracy of fuzzy templates of relative spatial relations. Further work is necessary to include more representative structures in different directions as reference structures in order to enhance the sensitivity of fuzzy templates of relative spatial relations at the expense of increased computation cost. Our method extracts tissue information from the test image and incorporate them into the prior knowledge to achieve high sensitive fuzzy maps for segmentation. But certain subcortical structures like thalamus has diffuse gray-level intensity distribution, which makes the tissue segmentation inaccurate. A more intelligent tissue segmentation could enhance the performance of segmentation [135]. Future work could focus on extracting additional useful information besides the tissue type information from the test image itself to account for each subject's own unique characteristics and subtle changes of shape. In the present approach, fuzzy templates are created based on T1-weighted images only. As multiple-channel images include more than one MRI property of tissues, it is possible to enhance the discriminating power of the method by incorporating new information, such as tissue or intensity fuzzy memberships from other multi-spectral or multi-modal images [136].

Chapter 4

Modelling Hemodynamic Variability of Brain Activation

Functional MRI measures brain activity *in vivo* both spatially and temporally. Previously, hypothesis-based and data-driven methods have been proposed to detect functional activation from fMRI. However, the complex nature and individual variability of hemodynamic response has not been properly addressed. The hemodynamic response function (HRF) models the coupling between the hemodynamic response and neuronal activity. In this chapter, we propose two fuzzy approaches: fuzzy feature modelling (FFM) and weighted contextual fuzzy c-means clustering (wcFCM), to detect brain activation accounting for individual variability of HRF. Both approaches are based on automatic extraction of a novel feature space from fMRI time-series using a sequence of temporal-sliding-windows (TSWs). Due to their capability of handling the variability among subjects, both methods outperform earlier SPM approach on group studies with synthetic fMRI data contaminated by both independent and correlated noise. Results also show that wcFCM is capable of incorporating contextual information of brain activity adaptively. Activation detected by the present approaches on real fMRI data were similar to those obtained with SPM. In the following sections, we first discuss the previous work in the detection of brain activation from fMRI and states the motivations to the proposed methods. Two fuzzy techniques are thereafter

described in detail, which is followed by experiments on both synthetic and real fMRI data. We discuss on the advantages and disadvantages of the methods in comparison with existing methods.

4.1 Introduction

The BOLD signal of fMRI is able to reveal brain activity both spatially and temporally. However, the detection of fMRI signal is not a trivial process as the BOLD signal change due to a typical experimental stimulation of the brain is very subtle, ranging from 1 to 5% on a 1.5 T scanner [137]. Furthermore, various noise and artifacts such as motion, electronic, physical, and physiological processes significantly confound fMRI signal. Therefore, the analysis techniques should be insensitive to uncertainties and fuzziness of fMRI data resulted from artifacts and noise.

Statistical parametric mapping (SPM) is the most widely used hypothesis-driven method for fMRI analysis, which assumes a linear regression model for fMR signal with a specific noise structure. The voxel-based method tests the hypothesis about fMR time-series by construction and assessment of spatially extended statistical processes based on Gaussian Random Field (GRF) theory [88]. However, in reality, the relationship between the variation in the fMRI signal and the stimulus presentation is non-linear [110]. And the hemodynamic response function (HRF) varies spatially, temporally, and among subjects [109]. Moreover, the structure of noise in fMRI is not well understood and remains a contentious subject [138]. Thus, the validity of the statistical models depends heavily on the extent to which the data satisfies the underlying assumptions.

In contrast, data-driven methods do not require any prior knowledge of the hemodynamic behavior and are considered more powerful and relevant for fMRI studies

when unknown or complex differential brain responses are expected [6]. The data-driven methods can be broadly divided into two groups: transformation-based and clustering-based methods. Principle component analysis (PCA) [139] and independent component analysis (ICA) belong to the first group, which transform the original high-dimensional input signal space to a space with components enabling the separation between functional responses and various other task-unrelated signals. ICA is an information theoretic approach separating the inputs into underlying signals or independent sources assuming a linear mixture. It has been applied both spatially [100] and temporally [101] to fMRI analysis. The clustering approaches consist of fuzzy clustering [105] and self-organizing map [9], attempting to classify time signals of the brain into several patterns according to temporal similarity. For these data-driven methods, usually the contents of one or more components are interpreted as activations because of its neurophysiologically feasible pattern. But the identification the classes is a difficult task and comprehension of other classes is often unclear.

Our objective is to develop a new approach that is capable of detecting functional activation from fMRI data, which is less vulnerable to noise and HRF variability than hypothesis-driven methods and is more interpretable than previous data-driven method. Two approaches based on fuzzy logic were developed in this thesis: fuzzy feature modelling (FFM) and weighted contextual fuzzy c-means clustering (wcFCM).

4.2 Fuzzy Feature Modelling

We propose a novel fuzzy feature modelling (FFM) method to detect activated voxels in the human brain, accounting for the individual variability of HRF. Five discriminating features were automatically extracted from fMRI by using a sequence of temporal-sliding-windows (TSW). The fuzzy model based on these features was first trained by using a gradient descent technique on an initial set of training voxels. The resulting

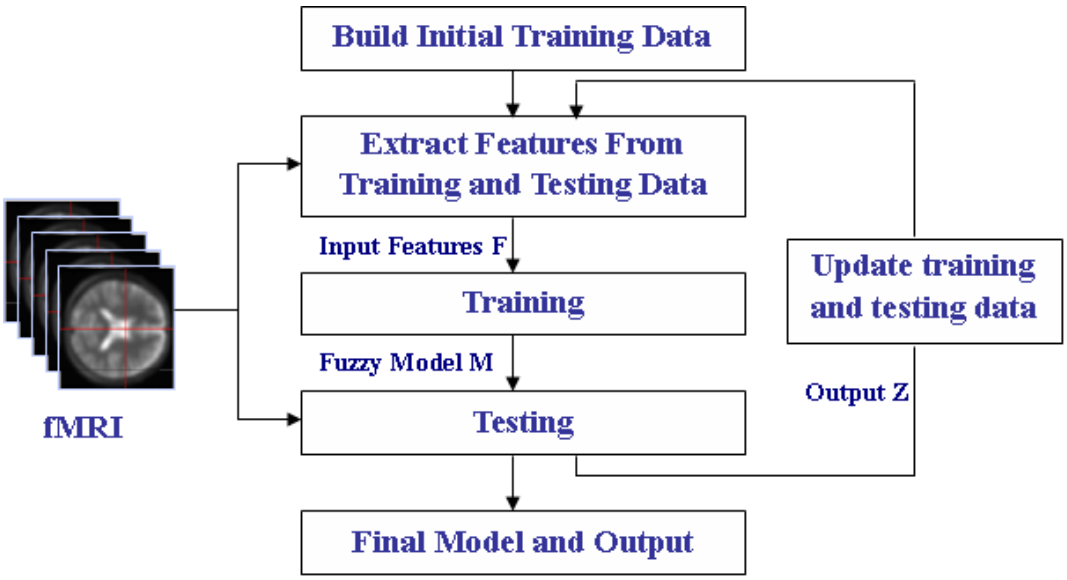


Figure 4.1: A schematic diagram illustrating the proposed fuzzy feature modelling approach for activation detection from fMRI.

fuzzy activation maps were incrementally updated to provide a measure of strength of activation of each voxel in the human brain. A two-way thresholding scheme was used to accurately determine activated voxels. A schematic diagram illustrating this approach is shown in Fig. 4.1.

Because of complex, nonlinear, and imprecise nature of fMRI signals and the presence of noise, accurate detection of brain activation from fMRI data with a classical linear or nonlinear modelling approach is difficult, especially in this case of the limited prior knowledge. In contrast, fuzzy modelling is a practical way to model fMRI data of high uncertainty when the information available is limited. Moreover, methods based on large volume of raw fMRI data often have high computational complexity and are vulnerable to noise. The proposed method of feature extraction is computationally less complex and has an increased ability to handle noise. In order to account for HRF variability across subjects, iterative learning of hemodynamic characteristics is capable of adapting the fuzzy model to individual subjects to better identify activated regions.

4.2.1 Extracting Features

Different brain voxels have different hemodynamic characteristics, for example, the magnitude of hemodynamic response of activated voxels and the time to reach the peak of hemodynamic response differ. The principle behind our feature extraction is a sequence of temporal-sliding-windows (TSW) shifted over a time-series, from which features that can consistently discriminate the activated and non-activated voxels of each condition are derived regardless of different shape, magnitude, or delay of hemodynamic response. Let $\Psi : \Omega \times \Theta \rightarrow Y$ be a functional time-series where $\Omega \subset \mathbf{N}^3$ denotes three-dimensional spatial domain of image voxels, $\Theta = \{1, 2, ...T\}$ index T number of 3D scans collected during an experiment, and $\times$ defines the Cartesian product of these two sets. Let $Y = \{y_{i,t} : i \in \Omega,\ t \in \Theta, y_{i,t} \in Q\}$ be a 4D data where Q denotes the range of image intensities and $y_{i,t}$ represents the intensity of voxel i at time t. Let Ω_B denote the set of brain voxels.

Here, we consider an experiment with one condition denoted by X for notational simplicity. The algorithm is also applicable for fMR time-series with a number of conditions. The condition X is presented together with a resting state alternatively or randomly for P times in a single run while T 3D brain scans are taken. Each block of condition X is denoted by B_p, $p = 1, 2..P$. Block B_p lasts for a duration of length l_p and the beginning of block B_p is denoted as b_p where $p = 1, 2, ...P$. This represents a general form of paradigms of both block and event-related designs in fMRI experiments. A sequence of temporal-sliding-windows (TSW) for the condition X is constructed from fMR time-series as follows:

1. Create a sequence of P number of windows, denoted by $W = (W_p : p = 1, 2, ...P)$; that is, one window W_p for each block B_p. The length of window W_p is denoted by w_p and suppose $w_p = l_p$, i.e., equals to the duration l_p of each block B_p. The initial starting point of window W_p is thus given by b_p.

2. Shift the sequence of windows W temporally forward by a sliding time interval s simultaneously. This results in a new sequence of windows $W(s) = (W_p(s) : s = 0, 1, ...S, p = 1, 2, ...P)$. Depending on different inter-scan time denoted by RT, the maximum sliding time interval $S = 32/\text{RT}$ varies based on the fact that the total length of hemodynamic response function is approximately 32 seconds. Thus, the starting and ending time of window $W_p(s)$ is $b_p + s$ and $b_p + s + w_p - 1$, denoted by $T_{p,1}(s)$ and $T_{p,2}(s)$ respectively for simplicity reason.

3. For each window $W_p(s)$ as $s = 0, 1, ...S$, the average intensity $A_p(i, s)$ of each voxel i is calculated as:

$$A_p(i,s) = \frac{\sum_{t=T_{p,1}(s)}^{\tilde{T}_{p,2}(s)} y_{i,t}}{\tilde{T}_{p,2}(s) - T_{p,1}(s)} \tag{4.1}$$

where $\tilde{T}_{p,2}(s) = \min\{T, T_{p,2}(s)\}$.

Thus, the total $S + 1$ sequences of P temporal-sliding-windows are able to capture the fMRI signal throughout the cycle of HRF in every scan continuously. We can observe a curve $A_p(i) = (A_p(i, s) : s = 0, 1, ...S)$ for each voxel i in each block B_p, whose shape discriminates the activated and non-activated voxels. Let us refer it as *quasi-hemodynamic curve* (QHC) due to its close relation with HRF because the QHC of activated class is largely depending on the HRF. Variations of HRF reflect in the change of shape of QHC $A_p(i)$ for each block B_p. Five discriminating features $f_k^p(i)$, $k = 1, 2, ...5$, are extracted from the QHC for each voxel i in each block B_p as follows:

1. Ratio of area under curve for QHC:

$$f_1^p(i) = \frac{\sum_{s=0}^{\tilde{w}_p} A_p(i,s)}{(\max_s A_p(i,s) - \min_s A_p(i,s)) \cdot \tilde{w}_p} \tag{4.2}$$

where $\tilde{w}_p = \min\{w_p, S\}$.

2. Ratio of area difference for QHC:

$$f_2^p(i) = \frac{\sum_{s=0}^{\tilde{w}_p} A_p(i,s)}{\sum_{s=\tilde{w}_p+1}^{S} A_p(i,s)} \tag{4.3}$$

3. Correlation between QHC $A_p(i)$ and the standard QHC, SA_p:

$$f_3^p(i) = \frac{\sum_{s=0}^{\tilde{w}_p}(A_p(i,s) - \overline{A_p(i,s)})(SA_p - \overline{SA_p})}{\sqrt{\sum_{s=0}^{\tilde{w}_p}(A_p(i,s) - \overline{A_p(i,s)})^2 \sum_{s=0}^{\tilde{w}_p}(SA_p - \overline{SA_p})^2}} \tag{4.4}$$

where $SA_p = \{SA_p(s) = -(s - \tilde{w}_p/2)^2 : s = 0, 1, ...\tilde{w}_p\}$.

4. Time ratio at peak amplitudes of QHC:

$$f_4^p(i) = \arg_{s\in[0,\tilde{w}_p]} \max_s A_p(i,s)/\tilde{w}_p \tag{4.5}$$

5. Time ratio at lowest amplitude for QHC:

$$f_5^p(i) = \arg_{s\in[0,\tilde{w}_p]} \min_s A_p(i,s)/\tilde{w}_p \tag{4.6}$$

Two curves are normalized to $[0, 1]$ before computation of correlation in feature 3 for fair comparison among voxels. The above five features extracted from QHC are capable of significantly discriminate activated and non-activated voxels.

The five features of each block are summed up over all the blocks as in Eq. 4.7. This averaging operation leads to a robust 5D feature space that is less vulnerable to noise or variations of HRF. It is still able to discriminate between activated or nonactivated voxels.

$$f_k(i) = \sum_{p=1}^{P} f_k^p(i)/P \tag{4.7}$$

Note that in our definitions, the duration of each block $l_p \leq S$ is assumed. For cases like $l_p > S$, we could apply the opposite settings, i.e., use window length $w_p = 32/\text{RT}$

and the maximum sliding time for each window $S_p = l_p$. The properties of the resulting curve is similar to QHC and same features could be extracted.

4.2.2 Learning Fuzzy Activation Maps

Based on the feature space developed in the previous section, we use a gradient descent technique to derive a fuzzy feature model. An incremental learning scheme is proposed to extract the useful knowledge not only from a set of canonical HRF characterized by different parameter values but also from fMRI dataset of each subject. This scheme is able to adapt to the hemodynamic variability among subjects by building one fuzzy model for each subject. The resulting fuzzy model provides the strength of activation of each voxel to each condition as well as a rule-base for interpretation of activation patterns. For each condition X, based on a vector of five input features, $\mathbf{f}(i) = (f_k(i) : k = 1, 2, ...5)$, a fuzzy model M is developed to give a single-output $Z(i)$ of activation strengths for all voxels $i \in \Omega_B$. The proposed incremental learning scheme to derive the model M is described:

1. *Building an initial set of training data*: A set of H different variations of original HRF are generated by randomly assigning the parameters (e.g., the delay and dispersion parameters) with values within a certain range. A set of time-series for the class of activated voxels are then created from input stimulus by modulating with HRF. Add all these time-series to the initial training set with the desired output D as 1. The initial set of training time-series for the class of non-activated voxels contains time-series of constant amplitude in the range $[0, 1]$ with the desired output D as 0. The initial set of test data includes the time-series of all brain voxels in the target subject, denoted by $Y = \{y_{i,t} : i \in \Omega_B, t \in \Theta\}$.

2. *Extracting features*: For each iteration of incremental, extract features from the

training data as described in section 4.2.1.

3. *Training*: Apply a gradient descent technique to train the fuzzy model M based on the extracted features until the convergence criterions are met. The objective of training is to minimize the error between the predicted output value $Z(i)$ and the desired output value $D(i)$ for each voxel i in the training set:

$$\epsilon = \frac{1}{2}(Z(i) - D(i))^2 \tag{4.8}$$

Gaussian fuzzy membership functions are presumed for input features:

$$\mu_k(i) = \exp(-\frac{1}{2}(\frac{f_k(i) - a_k}{\sigma_k})^2) \tag{4.9}$$

The mechanism of fuzzy model M employs a product t-norm for the input features (premise of the rule base) and a product implication. Delta output fuzzy membership function is used:

$$Z(i) = \frac{\sum_{c=1}^{2} b_c \prod_{k=1}^{5} \mu_k(i)}{\sum_{c=1}^{2} \prod_{k=1}^{5} \mu_k(i)} \tag{4.10}$$

where $c = 1$ and $c = 2$ denote the activated and non-activated classes of voxels, respectively. Therefore, parameters of the fuzzy model include input center a_k and variance σ_k of Gaussian fuzzy membership function and output center b_c of each class, where $c \in \{1, 2\}$ and $k = 1, 2, ...5$. The gradient descent technique is able to tune these parameters of the fuzzy model and detailed parameter updating equations can be found in [140].

4. *Testing*: Feed the extracted features of the testing data at the current round to the resulting fuzzy model M developed in step 3 to determine the output.

5. *Updating training data*: Select the most m activated and non-activated voxels in terms of highest or lowest outputs from the testing data. Add them into the

training set of next round. The desired output is again 1 for activated voxels and 0 for non-activated voxels. Meanwhile, remove the above selected points from the set of testing data. Repeat steps 2 to 5 till certain criteria is satisfied, e.g., the maximum iteration of incremental learning is reached or no more suitable voxels can be added into the set of training data.

6. *Final testing*: Apply the whole fMRI data to the developed fuzzy model to obtain the final output $Z = \{Z(i) : i \in \Omega_B\}$.

4.2.3 Detecting Activation

In order to detect activation by thresholding, a half-triangular fuzzy membership function is used to convert the original activation strength map $Z(i)$ to a fuzzy activation map $Z^* = \{Z^*(i) : i \in \Omega_B, Z^*(i) \in [0, 1]\}$:

$$Z^*(i) = \begin{cases} 1 & \text{if } Z(i) >= b_1 \\ (Z(i) - b_2)/(b_1 - b_2) & \text{if } Z(i) < b_1 \ \& \ Z(i) > b_2 \\ 0 & \text{if } Z(i) <= b_2 \end{cases} \tag{4.11}$$

This map represents the activation strength of each voxel in terms of fuzzy membership values, which enables a rather consistent and effective performance among differentially responded voxels, with the following two-way thresholding scheme.

First, a simple thresholding is applied based on two parameters ζ_1 and ζ_2 ($\zeta_1 > \zeta_2$) to add high confident voxels to a set of activated voxels Υ_1 and a set of non-activated voxels Υ_2:

$$\beta(i) = \begin{cases} 1 & \text{if } Z^*(i) >= \zeta_1 \\ 0 & \text{if } Z^*(i) <= \zeta_2 \end{cases} \tag{4.12}$$

Then, for those undetermined voxels with $Z^*(i) < \zeta_1$ and $Z^*(i) > \zeta_2$, an ordered list is formulated based on the value of $Z^*(i)$. A screening process begins simultaneously

from the two ends of this list and assigns voxels an activated or non-activated label as follows:

$$\beta(i) = \begin{cases} 1 & \text{if } \chi(1,i,v) > \chi(0,i,v) \\ 0 & \text{if } \chi(1,i,v) < \chi(0,i,v) \end{cases} \tag{4.13}$$

where $\chi(e,i,v) = \prod_{j\in\varphi(i,v), j\in\Upsilon_e, \beta(j)=e} |Z^*(j) - Z^*(i)| \cdot \widehat{distance}(i,j)$ and $e = \{0,1\}$. $\varphi(i,v)$ is the neighborhood of voxel i within a cube of size v^3. Each voxel i is removed from list and added into the set of activated voxels Υ_1 or the set of non-activated voxels Υ_2 accordingly. The screening process continues until there are no more voxels left in the list.

In order to detect the common set of activated brain regions in a group study of N subjects, fusion of fuzzy activation maps Z^* of each subject is performed to form a combined fuzzy activation map $Z^*_{group} = \{Z^*_{group}(i) : i \in \Omega_B, Z^*_{group}(i) \in [0,1]\}$:

$$Z^*_{group}(i) = (\prod_{j=1}^{N} Z^*_j(i))^{1/N} \tag{4.14}$$

Note that the activation map is normalized to $[0,1]$ after fusion and the same thresholding procedure is applied to the combined fuzzy activation map to identify brain activation.

4.3 FCM with Contextual Constraints

The activation of each brain voxel is affected by the contextual information of the neighboring voxels which often have similar response characteristics to input stimuli. Gaussian smoothing is often applied to enhance signal to noise ratio (SNR) before statistical analysis, implicitly accounting for spatial dependency. This leads to overly smoothed images and a loss of high-frequency information. Markov random fields (MRF) [141, 142, 106, 143] and conditional random fields (CRF) [107] have been

attempted to explicitly incorporate spatial and temporal correlation for the detection of brain activation from fMRI data. A MRF based on mixture models has been proposed for fMRI segmentation with adaptive spatial regularization [144]. Autoregressive spatio-temporal models have also been used to incorporate tissue-type related noise priors for spatial constraints in fMRI analysis [109]. The above approaches still suffer from the assumptions on structures of HRF and noise.

Fuzzy c-means clustering (FCM) has been widely used in image segmentation, which allows incorporation of fuzziness to the belongingness of a voxel to a particular class or an object. The contextual information is handled by a contextual regularized term in the cost function of FCM, for example, smooth constraints on the bias field of intensity inhomogeneity correction [145] and a term for immediate neighborhood information for tissue segmentation of anatomical MR images [146]. An unsupervised tissue segmentation of structural MRI by using a penalized fuzzy clustering algorithm has also been proposed [147]. They incorporate spatial neighborhood information by using a penalty term. The kernelled version of FCM with spatial constraints has been recently applied to further enhance image segmentation accuracy [148].

Based on the same discriminating feature space in FFM, an unsupervised weighted contextual fuzzy c-means clustering (wcFCM) is proposed to identify activated voxels. The cost function of FCM, regularized by spatial context, is adopted to capture local and neighboring information. FCM with contextual modelling on novel fuzzy feature space is similar to kernelized FCM on original fMR time-series. However, the advantage of our approach is that the kernel is well defined by the new feature space compared to approximating with a normal Gaussian or polynomial function. In addition, the weighted regularized term for each neighboring voxel enables incorporation of contextual constraints adaptively, i.e., the neighboring voxels with similar properties as reference voxel play a greater role than dissimilar ones in computing the fuzzy memberships.

There are two major steps in the proposed wcFCM approach: (1) extracting the discriminating features from fMRI time-series using a sequence of temporal-sliding-windows (TSWs) as described in section 4.2.1 and (2) the use of wcFCM to classify voxels of each individual subject into activated and non-activated classes. This scheme is able to adapt to the hemodynamic variability of each subject. The extraction of features from fMRI time-series reduces computational complexity of the analysis and enhances the capability of handling noise. wcFCM is able to provide (1) the strength of activation of each voxel to each condition, (2) a clear classification of voxels by comparing membership values, and (3) a rule-base for interpretation of activation patterns. In what follows, the details of the present approach is described and the performance of wcFCM is illustrated with experiments on both individual and group study of synthetic and real fMRI data.

4.3.1 Contextual Fuzzy *C*-means Clustering (cFCM)

For a given condition X, each voxel $i \in \Omega$ has a corresponding vector of five features $\mathbf{f}(i) = \{f_k(i) : k = 1, 2, ...5\}$. We use FCM on these features to classify voxels into two classes: activated or non-activated. The objective function of standard FCM partitions the feature set $\{\mathbf{f}(i) : i \in \Omega\}$ into C clusters and is given by

$$J_m = \sum_{c=1}^{C} \sum_{i \in \Omega} u^m(c, i) ||\mathbf{f}(i) - \mathbf{v}(c)||^2 \tag{4.15}$$

subject to constraints

$$u(c, i) \in [0, 1], \quad \sum_{c=1}^{C} u(c, i) = 1, \forall i \text{ and } 0 < \sum_{i \in \Omega} u(c, i) < |\Omega|, \forall c \tag{4.16}$$

where $u(c, i)$ is the fuzzy membership value of voxel i belonging to class c, $\mathbf{v}(c)$ is the feature vector of the centroid of class c, $||\cdot||$ stands for Euclidean norm, and parameter

m is the weighting exponent of fuzzy memberships.

Standard FCM fails to segment images corrupted by noise, outliers, and other artifacts [148]. The robustness of FCM to noise is improved by directly modifying the objective function of Eq. 4.15 in order to incorporate contextual information [146]:

$$\begin{aligned} J_m &= \sum_{c=1}^{C}\sum_{i\in\Omega} u^m(c,i)||\mathbf{f}(i)-\mathbf{v}(c)||^2 + \\ &\quad \frac{\alpha}{|Ne(i)|}\sum_{c=1}^{C}\sum_{i\in\Omega} u^m(c,i)\sum_{r\in Ne(i)}||\mathbf{f}(r)-\mathbf{v}(c)||^2 \end{aligned} \tag{4.17}$$

where $Ne(i)$ stands for the set of neighbors surrounding voxel i and $|Ne(i)|$ is the number of valid voxels in the neighborhood. The parameter α controls the effect of regularization from neighboring voxels. This regulariazation term is minimized when neighboring voxels of target voxel are also closer to the centroid. The clusters are obtained by solving

$$\text{minimize} \quad J_m, \tag{4.18}$$

$$\text{subject to} \quad \sum_{c=1}^{C} u(c,i) = 1, \forall i \tag{4.19}$$

With the above minimization, higher membership values are given to voxels closer to the centroid and lower membership values are given to those further from the centroid. The Lagrange method is adopted to incorporate the constraints into the optimization and the augmented object function is given by:

$$J_m{}' = J_m + \lambda(1 - \sum_{c=1}^{C} u(c,i)) \tag{4.20}$$

This problem is solved by taking the derivatives of $J_m^{'}$ with $u(c,i)$ and applying the

constraints. The necessary conditions for the local minimum are

$$u(c,i) = \frac{(||\mathbf{f}(i) - \mathbf{v}(c)||^2 + \frac{\alpha}{|Ne(i)|}\sum_{r\in Ne(i)}||\mathbf{f}(r) - \mathbf{v}(c)||^2)^{\frac{-1}{m-1}}}{\sum_{c=1}^{C}(||\mathbf{f}(i) - \mathbf{v}(c)||^2 + \frac{\alpha}{|Ne(i)|}\sum_{r\in Ne(i)}||\mathbf{f}(r) - \mathbf{v}(c)||^2)^{\frac{-1}{m-1}}} \tag{4.21}$$

$$\mathbf{v}(c) = \frac{\sum_{i\in\Omega} u^m(c,i)(\mathbf{f}(i) + \frac{\alpha}{|Ne(i)|}\sum_{r\in Ne(i)}\mathbf{f}(r))}{(1+\alpha)\sum_{i\in\Omega} u^m(c,i)} \tag{4.22}$$

4.3.2 Weighted Contextual Fuzzy *C*-means Clustering (wcFCM)

The above contextual fuzzy c-means clustering approach considers the relationship between neighboring voxels and cluster centroids for minimization. However, the relationship between neighboring voxels and the reference voxel is not taken into account. Neighboring voxels that are more similar to the reference voxel in terms of extracted properties should be given more weight than those less similar ones. Therefore, we modify the cost function in cFCM (Eq. 4.17) to incorporate the contextual information by weighting with the similarity to the reference voxel. We refer to this method as weighted contextual fuzzy c-means clustering (wcFCM). The modified cost function is given by:

$$\begin{aligned} J_m &= \sum_{c=1}^{C}\sum_{i\in\Omega} u^m(c,i)||\mathbf{f}(i) - \mathbf{v}(c)||^2 + \\ &\quad \frac{\alpha}{|Ne(i)|}\sum_{c=1}^{C}\sum_{i\in\Omega} u^m(c,i)\sum_{r\in Ne(i)}\frac{||\mathbf{f}(r) - \mathbf{v}(c)||^2}{||\mathbf{f}(r) - \mathbf{f}(i)||^2} \end{aligned} \tag{4.23}$$

where $Ne(i)$ stands for the set of neighbors around voxels i and $|Ne(i)|$ is the number of valid voxels in the neighborhood. Similarly, the clusters are obtained by solving the Eq. 4.18 by the Lagrange method. The necessary conditions for the local minimum

in the proposed weighted contextual fuzzy c-means clustering approach are:

$$u(c,i) = \frac{(||\mathbf{f}(i) - \mathbf{v}(c)||^2 + \frac{\alpha}{|Ne(i)|}\sum_{r\in Ne(i)} \frac{||\mathbf{f}(r)-\mathbf{v}(c)||^2}{||\mathbf{f}(r)-\mathbf{f}(i)||^2})^{\frac{-1}{m-1}}}{\sum_{c=1}^{C}(||\mathbf{f}(i) - \mathbf{v}(c)||^2 + \frac{\alpha}{|Ne(i)|}\sum_{r\in Ne(i)} \frac{||\mathbf{f}(r)-\mathbf{v}(c)||^2}{||\mathbf{f}(r)-\mathbf{f}(i)||^2})^{\frac{-1}{m-1}}} \tag{4.24}$$

$$\mathbf{v}(c) = \frac{\sum_{i\in\Omega} u^m(c,i)(\mathbf{f}(i) + \frac{\alpha}{|Ne(i)|}\sum_{r\in Ne(i)} \frac{\mathbf{f}(r)}{||\mathbf{f}(r)-\mathbf{f}(i)||^2})}{\sum_{i\in\Omega} u^m(c,i)(1 + \frac{\alpha}{|Ne(i)|}\sum_{r\in Ne(i)} \frac{1}{||\mathbf{f}(r)-\mathbf{f}(i)||^2})} \tag{4.25}$$

4.3.3 Convergence of cFCM and wcFCM

Intuitively, the objective function in Eq. 4.23 reduces by both steps corresponding to Eqs. 4.21 and 4.22, respectively, so the combined procedure makes the function descent strictly. In this section, a mathematical analysis of the convergence condition of the proposed algorithms is presented. For cFCM, Eq. 4.22 can be rewritten as

$$(1+\alpha)\sum_{i\in\Omega} u^m(c,i)\mathbf{v}(c) = \sum_{i\in\Omega} u^m(c,i)(\mathbf{f}(i) + \frac{\alpha}{|Ne(i)|}\sum_{r\in Ne(i)} \mathbf{f}(r)) \tag{4.26}$$

which represents a set of linear equations in terms of $\mathbf{v}(c)$ if $u(c,i)$ is constant. The convergence property of the cFCM algorithm is to view Eq. 4.22 as the Gauss-Seidel iteration for solving the set of linear equations. The Gauss-Seidel algorithm is guaranteed to converge if the matrix representing the equations is diagonally dominant [149, 150]. Equation 4.26 goes from $c = 1$ to C and each $\mathbf{v}(c)$ is k-dimensional ($k = 5$), that is $\mathbf{v}(c) = \begin{bmatrix} v_1(c) & v_2(c) & \dots & v_k(c) \end{bmatrix}^T$.

Thus, the corresponding matrix has the size of $Ck \times Ck$ and has the block diagonal form as Eqs. 4.27-4.28.

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & 0 & . & . & 0 \\ 0 & \mathbf{A}_2 & & & \\ . & & . & & \\ . & & & . & \\ 0 & & & & \mathbf{A}_k \end{pmatrix} \tag{4.27}$$

where

$$\mathbf{A}_1 = \mathbf{A}_2 = ... = \mathbf{A}_k = \begin{pmatrix} (1+\alpha)q_1 & 0 & . & . & . & 0 \\ 0 & (1+\alpha)q_2 & & & & \\ . & & (1+\alpha)q_c & & & \\ . & & & & . & \\ 0 & & & & & (1+\alpha)q_C \end{pmatrix} \tag{4.28}$$

where $q_c = \sum_{i \in \Omega} u^m(c,i)$. It is observed that the matrix representing the equations is the Hessian (second order derivative) of J_m with respect to $\mathbf{v}(c)$, which is diagonal matrix and is positive definite. From the Gauss-Seidel iteration point of view, the matrix is also strict row diagonal dominance, meaning that for each row, the absolute value of the diagonal term is greater than the sum of absolute values of other terms. Therefore, the iteration would converge if Eq. 4.22 is applied repetitively with $u(c,i)$ kept constant. In practice, both Eq. 4.21 and 4.22 are applied alternatively in the iterations. The Hessian of J_m with respect to $u(c,i)$ is diagonal matrix and is positive definite. Although the joint Hessian matrix of J_m with respect to both $u(c,i)$ and $\mathbf{v}(c)$ have nonzero off-diagonal elements and its eigenvalues cannot be easily determined. According to the theorems derived by Bezdek [151, 152], it is concluded that the cFCM algorithm converges, at least along a subsequence, to a local optimal solution. The reformulation of Eq. 4.25 and convergence proof for wcFCM can be derived in

the similar manner.

4.3.4 Algorithms of cFCM and wcFCM

The iterative learning procedure used in cFCM and wcFCM could be summarized as follows:

Begin

Set the initial centroids $\mathbf{v}(c)$ of C clusters, $c = 1, 2, ...C$, and choose $\varepsilon(> 0)$ to a very small value

repeat

Calculate the fuzzy membership values $u(c, i)$ of each voxel i to all clusters by Eq. 4.21 for cFCM and Eq. 4.24 for wcFCM

Calculate the centroids $\mathbf{v}(c)$ based on the updated $u(c, i)$ by Eq. 4.22 for cFCM and Eq. 4.25 for wcFCM

until $\xi < \varepsilon$

End

The average difference ξ between centroids obtained in consecutive updating rounds based on the feature space is defined as:

$$\xi = \left(\frac{\sum_{c=1}^{C} ||\mathbf{v}_{prev}(c) - \mathbf{v}_{curr}(c)||^2}{C * K}\right)^{0.5} \tag{4.29}$$

where K represents the number of features. The final fuzzy membership maps of both activated and non-activated classes provide a concrete segmentation of activated and non-activated brain regions from fMRI data. The voxel is assigned to the class with the highest fuzzy membership value.

4.4 Experiments on Synthetic Data

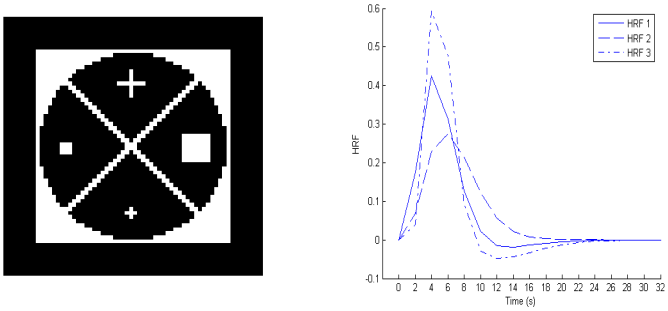
All simulations were done in MATLAB. Both synthetic and real fMRI data were used in the experiments and a comparison between the results produced by the proposed approaches and statistical parametric mapping (SPM) is given.

4.4.1 Data Generation

A 2D synthetic functional dataset consisting of six cycles (with each cycle having 8 ON and 8 OFF scans, TR= 2s, $T = 96$) was simulated. The responses of the activated voxels were generated by convolving a box-car time-series with the HRF, using a mixture of two gamma functions, while the non-activated voxels were kept unchanged at the value of zero. The ground truth of activation pattern is shown in Fig. 4.2a. Different levels of independent Gaussian noise and spatially correlated noise (by averaging the neighboring i.i.d. Gaussian noise) were added into the synthetic data: signal to noise ratio (SNR)=$\{2.0, 1.2, 0.45\}$. In order to evaluate the influence of inter-subject variability of HRF on system performance, five synthetic time-series were simulated based on different HRFs. The parameters of HRF include the delay of response and undershoot relative to onset, the dispersion of response and undershoot, the ratio of response to undershoot, and the total length of HRF function. Fig. 4.2b give examples of different HRFs by varying the values of its parameters.

4.4.2 Parameter Settings

The SPM analysis used the standard procedure implemented in the SPM2 toolbox [89]. A 3D Gaussian filter having FWHM= 6mm was applied to smooth the images to have the characteristics of GRF in order to fulfill the statistical assumptions of SPM as well as to increase the signal to noise ratio of the images. Canonical HRF and HRF



(a) Truth (b) HRFs

Figure 4.2: Generation of synthetic fMRI data: (a) true activated regions and (b) example of three HRFs for the synthetic time-series of different subjects.

with time and dispersion derivatives were both used as the basis function for statistical modelling. However, it was found that the latter achieved the same level of sensitivity with more false positives than the former. Therefore, the results reported here were based on canonical HRF. Confounding effects of global fluctuations of time-series were removed by proportional scaling. Low frequency noise was removed using a high pass filter with a cut-off frequency of 128s. The group analyses of five synthetic time-series were performed using the fixed-effect analysis (FFX) [153]. Specific effects were tested by applying appropriate linear contrasts and t-statistical parametric maps were used to assess significant hemodynamic changes. We report significant activations of voxels at a significant level of $p < 0.05$ corrected for multiple comparisons, using the false discovery rate (FDR) technique [97].

The proposed FFM approach was performed on original fMRI data without smoothing as no assumptions of data is required. Twenty time-series for activated and non-activated class were used, $H = 20$, as the initial set of training data. At each iteration of the incremental learning process, twenty voxels of the test images were added into the training data of each class. The value of H is selected such that the initial training data is large enough to make a differentiation between classes and the following updates recruit fewer confounding voxels for training. It is observed that more iter-

ations of incremental learning of FFM approach is capable of better adapting to the data but overfitting along with performance degration may occur when confounding voxels are added in the training process. The computational complexity also increases as the iteration continues. Thus, ten iterations of incremental learning were used in the experiments. For step 3 in section 4.2.2, the initial values of fuzzy membership parameters of all classes were set as follows: input center a_k equals to f_k of the first training sample, input variance $\sigma_k = 1$, output centers $b_1 = 1$ and $b_2 = 0$ for activated and nonactivated classes, respectively. Parameters controlling the updating steps for a, σ, and b were all initialized to 1. The training stops when the average square error ϵ defined in Eq. 4.8 is less than 10^{-3} or the absolute change of ϵ between consecutive iterations is less than 10^{-5}. In section 4.2.3, the neighborhood size of the undetermined voxels $v = 5$. If no neighbor with known class label was found, the search continues with a larger neighborhood by increasing v by 1. However, the neighborhood size was limited to 1/125 of the whole brain volume to prevent over-searching.

Both cFCM and wcFCM were performed on original (unsmoothed) time-series with $\alpha = 3$, $m = 2$, $\varepsilon = 0.001$. The optimal value of α was empirically determined. A six-voxels neighborhood was used for contextual modelling of the target voxel. Initial cluster centroids were set to the feature vector of voxels of the highest and lowest correlations with the input stimulus modulated by canonical HRF. Extracted features were individually scaled to $[0, 1]$ before applying cFCM and wcFCM such that each feature plays an equal weight.

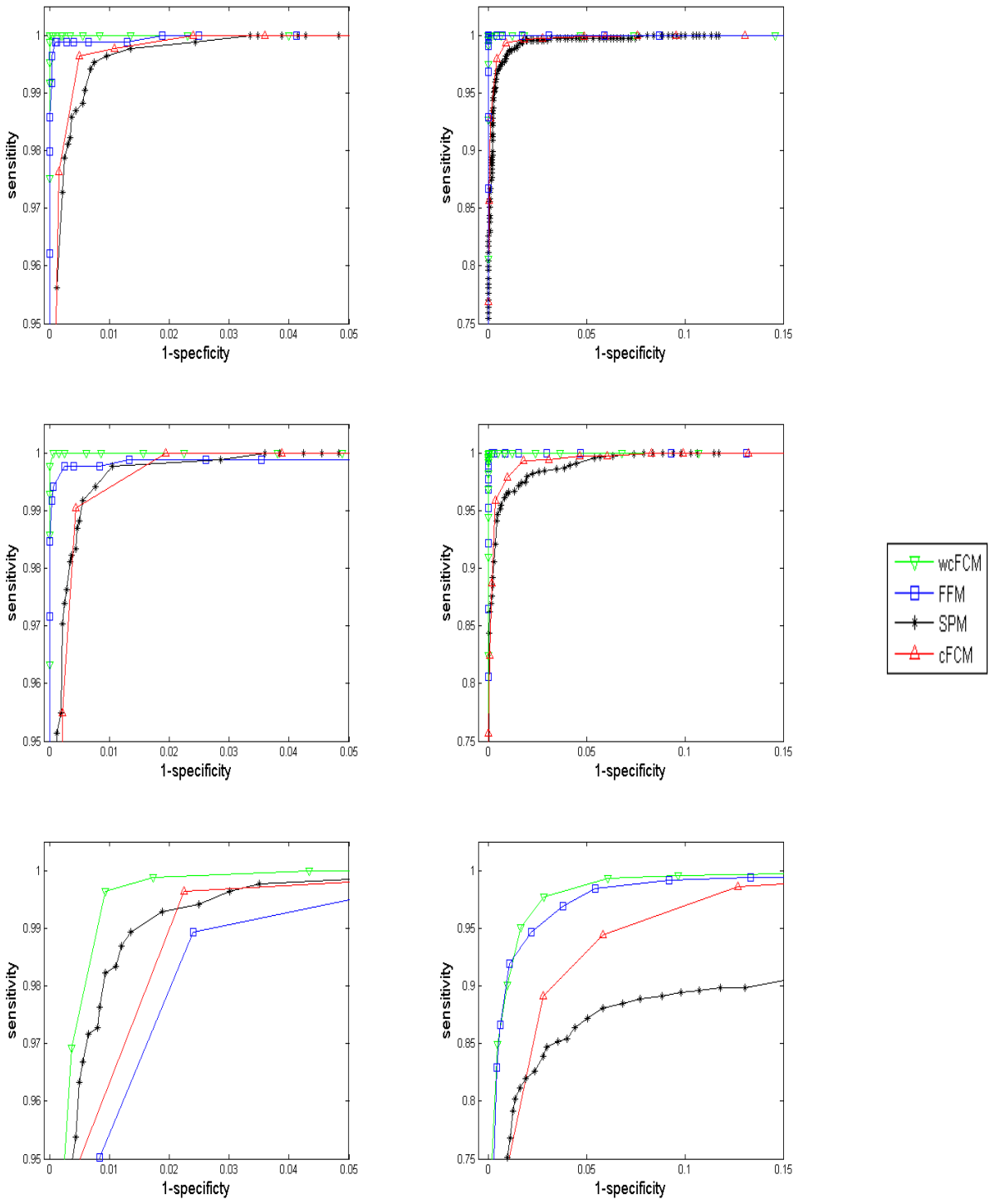
With the above settings, the execution time of SPM, FFM, cFCM, and wcFCM on the synthetic data were approximately 1.5min, 4min, 1.5min, and 4.5min, respectively, on a Pentium IV 2.4GHz machine having 512MB memory and running MATLAB version 7.0 on Windows XP environment.

4.4.3 Performance Evaluation

The performance of the detection of functional activation from a group of five synthetic fMRI data by SPM, FFM, cFCM, and wcFCM were compared. Fig. 4.3 show the ROC curves of all four approaches for synthetic fMRI data with Gaussian and correlated noise at three different levels of signal to noise ratios.

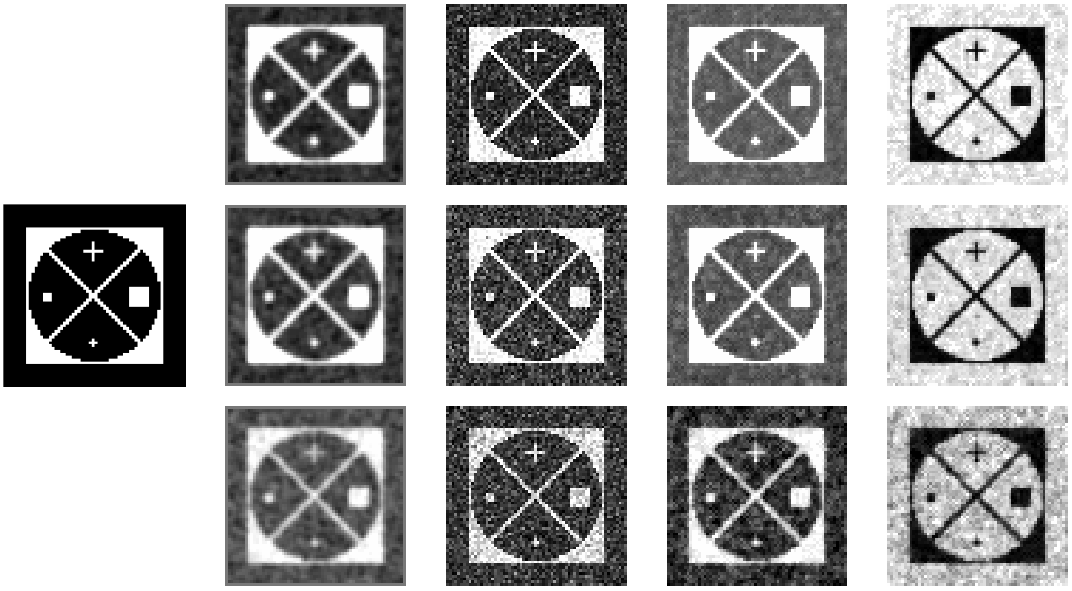
As seen, the weighted contextual FCM approach outperforms the other three approaches in all cases as it is able to cope with inter-subject variability in HRFs as well as to incorporate contextual information adaptively. The weighted contextual FCM has significant improvement over original contextual FCM (cFCM) for both correlated and Gaussian noises. FFM approach performs better than SPM and cFCM for data with correlated noise as it could adapt to HRF variability by incremental learning. However, FFM deteriorates fast under Gaussian noise because it does not take neighborhood information into account. Experiments also showed that SPM suffered more from correlated noise than Gaussian noise. SPM assumes a fixed structure of fMRI signals and noise for statistical analysis and therefore cannot handle the variability of HRF and high level of correlated noise. Moreover, the preprocessing step of smoothing in SPM may lose high-frequency information.

The final detected activation by SPM, FFM, cFCM, and wcFCM are shown in Fig. 4.6 for data with Gaussian noise and Fig. 4.7 for data with correlated noise. For SPM, threhsolding methods should be catered to correct for multiple comparisons. Familywise error rate (FWER) thresholding method [95, 96] usually produces more conservative results than false discovery rate (FDR) method [97]. In this experiment, we found that FDR was more appropriate for images with correlated noise while FWER was more appropriate for images with independent noise. Here, the results thresholded by FDR are reported. The two-way thresholding technique of FFM is effective, yet requires more computation. The elegance of wcFCM and cFCM is that



(a) Independent noises (b) Correlated noises

Figure 4.3: ROC curves of detecting activation on a group study of five synthetic fMRI data with different HRFs by SPM, FFM, cFCM, and wcFCM. Different levels of (a) independent Gaussian noise and (b) correlated noise were tested: SNR= 2.0 (row 1), SNR= 1.2 (row 2), and SNR= 0.45 (row 3).

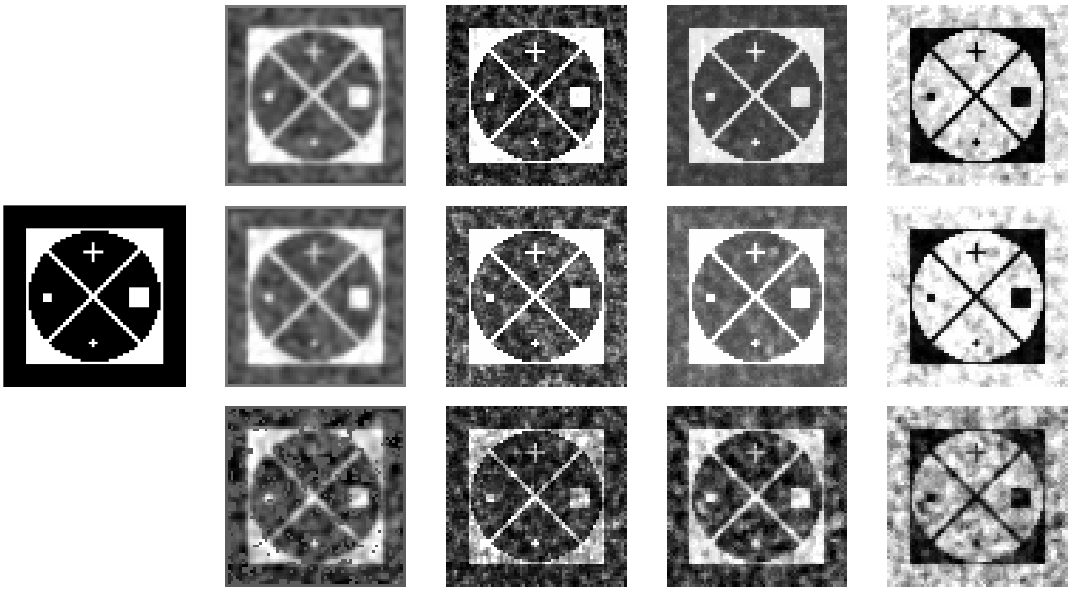


Truth (a) SPM (b) FFM (c) wcFCM1 (d) wcFCM2

Figure 4.4: Comparison of the unthresholded activation maps from synthetic fMRI data with Gaussian noise by SPM, FFM, and wcFCM. Column (c) and (d) denotes the fuzzy maps for activated and non-activated classes, respectively, in wcFCM approach. Different levels of noise were tested: SNR= 2.0 (row 1), SNR= 1.2 (row 2), and SNR= 0.45 (row 3).

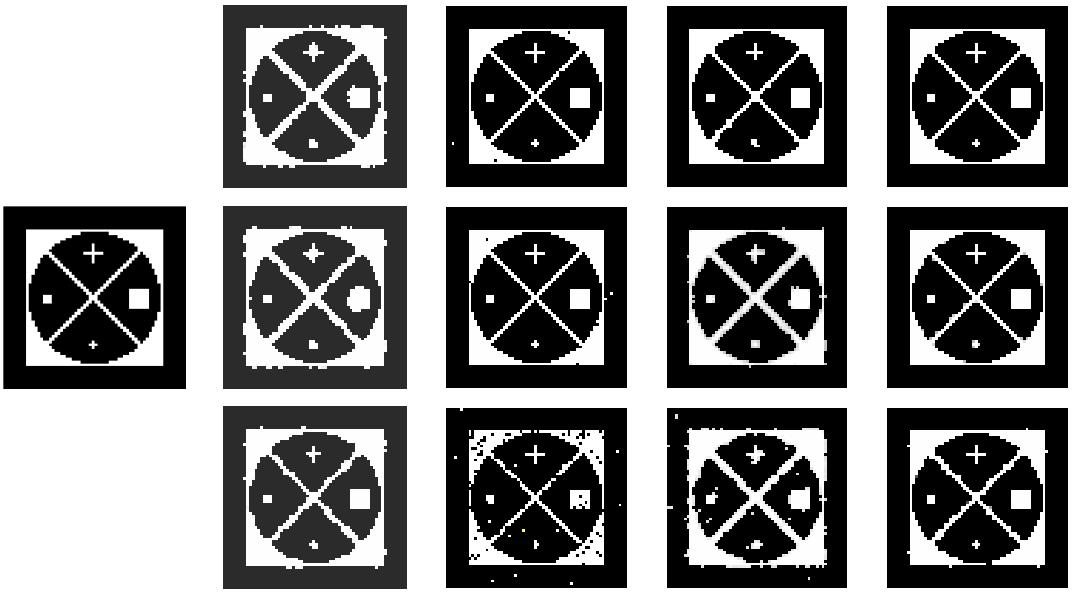
no thresholding is involved; activated voxels are automatically identified by comparing the fuzzy membership values of activated and non-activated classes, i.e., winner-takes-all rule. To further evaluate the differences between the proposed approaches and SPM, a direct comparison between the unthresholded statistical map of SPM and unthresholded fuzzy activation map of FFM, and wcFCM is shown in Fig. 4.4 and 4.5. Since wcFCM depends on the competition of two clases which does not require thresholding, fuzzy maps of both activated and nonactivated classes are shown as in column (c) and (d).

As seen in Fig. 4.6 and Fig. 4.7, wcFCM discovers more accurate and detailed activation than other approaches, especially in the case of high correlated noise. FFM does not incorporate neighborhood information and thus suffers from high level of noise as more false positive and false negative voxels appear. Statistical parametric mapping approach has rather good performance for data with Gaussian noise, which



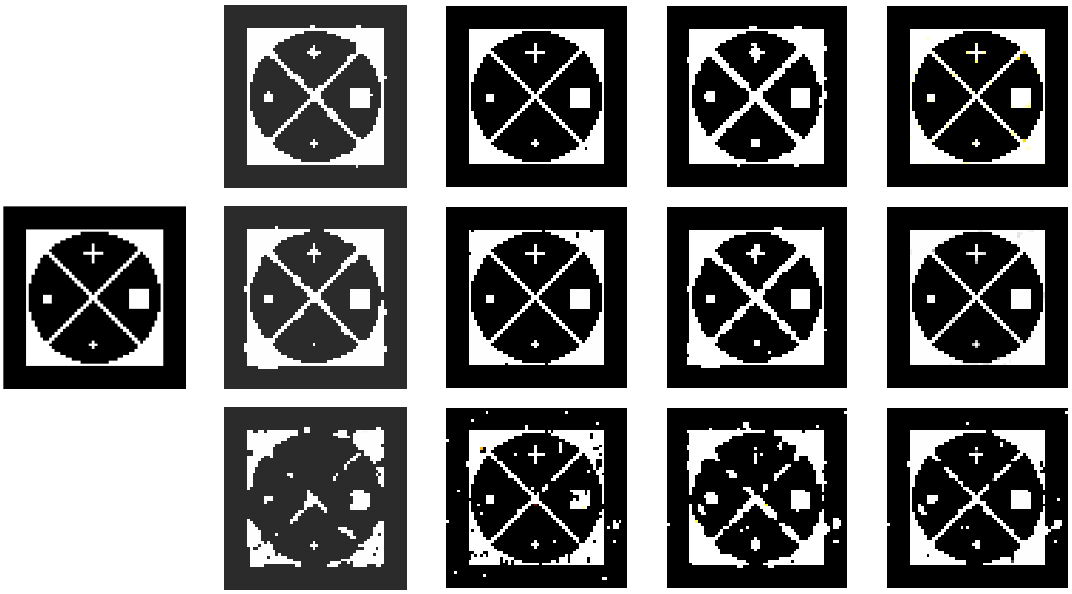
Truth (a) SPM (b) FFM (c) wcFCM1 (d) wcFCM2

Figure 4.5: Comparison of the unthresholded activation maps from synthetic fMRI data with correlated noise by SPM, FFM, and wcFCM. Column (c) and (d) denotes the fuzzy maps for activated and non-activated classes, respectively, in wcFCM approach. Different levels of noise were tested: SNR= 2.0 (row 1), SNR= 1.2 (row 2), and SNR= 0.45 (row 3).



Truth (a) SPM (b) FFM (c) cFCM (d) wcFCM

Figure 4.6: Comparison of detected activation from synthetic fMRI data with independent Gaussian noise by SPM, FFM, cFCM, and wcFCM. Different levels of noise were tested: SNR= 2.0 (row 1), SNR= 1.2 (row 2), and SNR= 0.45 (row 3).



Truth (a) SPM (b) FFM (c) cFCM (d) wcFCM

Figure 4.7: Comparison of detected activation from synthetic fMRI data with correlated noise by SPM, FFM, cFCM, and wcFCM. Different levels of noise were tested: SNR= 2.0 (row 1), SNR= 1.2 (row 2), and SNR= 0.45 (row 3).

satisfies its assumption about the nature of noise. However, SPM could not detect important activation for data with high level of correlated noise. Over-smoothing effect is also observed in SPM. By comparing the results of cFCM and wcFCM, it is observed that the adaptive manner of contextual modelling by considering the similarity between neighboring voxels and reference voxel significantly improves the performance, i.e., fewer false positive and false negative voxels exist.

Moreover, the feature extraction scheme in the proposed approaches is able to capture the variability of HRF, which could be proved by the resulting feature centroids of activated and non-activated classes in both FFM and FCM. Three different HRFs are shown in Fig. 4.2(b). The centroids of all five input features in the resulting three fuzzy models in FFM are given in table 4.1. By comparing the shapes of HRFs and the centroids of fuzzy models, we can see that the resulting feature centroids are highly related to the original parameters of HRFs used for the generation of synthetic fMRI data. Since the delay of response and undershoot for HRF-1 and -3 is smaller

HRF	f_1	f_2	f_3	f_4	f_5
1	0.81 (0.10)	2.86 (0.86)	0.51 (-0.08)	0.49 (0.11)	0.65 (0.02)
2	0.81 (0.14)	3.10 (0.89)	0.50 (-0.02)	0.58 (0.10)	0.80 (0.06)
3	0.82 (0.10)	2.97 (0.86)	0.54 (-0.08)	0.53 (0.09)	0.66 (0.01)

Table 4.1: Correlation between properties of HRF and parameters of fuzzy feature models. HRF number 1, 2, and 3 refers to the HRF curves in Fig. 4.2 (b). The numbers outside and inside the brackets are the centers of five features (defined in section 4.2.1) of the resulting fuzzy feature model for activated and non-activated class, respectively.

than HRF-2, the distance between the input centers of feature f_4 (time ratio at peak amplitude for QHC) and feature f_5 (time ratio at lowest amplitude for QHC) in activated and non-activated class are smaller in fuzzy model of HRF-1 and -3 than in fuzzy model of HRF-2. HRF-2 has little undershoot than the other two HRFs, and hence, the distance between input centers of feature f_2 (area difference ratio for QHC) in activated and non-activated class is larger in fuzzy model of HRF-2 than the other two. The feature extraction step is capable of modelling HRF properties accurately and different fuzzy models are built to account for the variability of HRF. As a result, the resulting centroids of each cluster reflect meaningful information of a certain cognitive task performed by individual subjects.

4.5 Experiments on Real Data

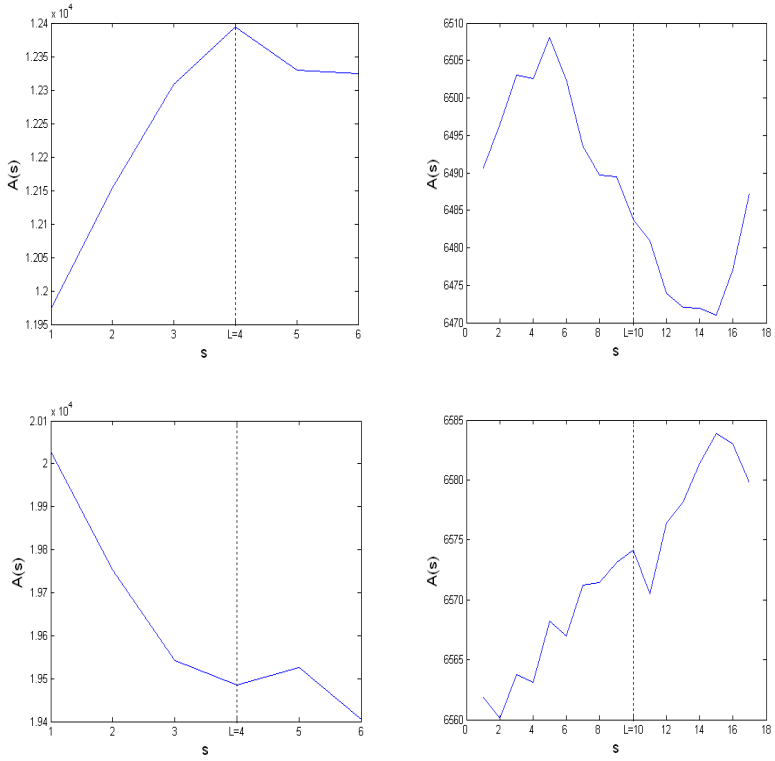
The proposed approaches are also demonstrated by three sets of real fMRI data performing: (1) visual task, (2) counting Stroop task, and (3) silent reading task. The approaches are validated by comparing the results with SPM and literature. The following three subsections describe the details of each dataset and present the corresponding results.

4.5.1 Visual Task

The detection approaches were tested on real fMRI data of visual stimulation task. For each subject, 64 brain scans with three 2D $T2^*$-weighted slices per image scan were acquired by using a gradient-echo fast low angle shot (FLASH) sequence. The matrix size is 128×128, TE/TR are 40ms/80.5ms, and voxel size is 1.953×1.953mm. The corresponding two-dimensional anatomical slices were also acquired with a T1-weighted IR RARE sequence (TE/TR are 3900ms/40ms; matrix size is $= 512 \times 512$) in the same experiment session. In the experiment, ON and OFF stimuli were presented at a rate of 5.16s per sample. Each stimulation period had four successive stimulation of ON samples followed by four stimulation of OFF samples. The stimulations were repeated for eight cycles (total experiment time = 5.5 min) and experiments were carried out at different sessions with different subjects. An 8-Hz alternating checker-board pattern with a central fixation point was projected on a LCD system, and the subject was asked to fixate on the point during stimulations. Further details of the visual simulation experiments can be found in [25].

For SPM analysis, all functional images were first corrected for movement artifacts, resampled, and smoothed with a 3D Gaussian filter having FWHM= 4.47mm. T-contrast is used for statistical analysis in SPM2 with canonical HRF as basis function. Voxels with $p < 0.05$ corrected using familywise error rate(FWER) is determined to be activated. The proposed FFM, cFCM, and wcFCM approaches were applied on the original unsmoothed images that have been corrected for movement artifacts with the same parameter settings used for synthetic data.

In order to illustrate the usefulness of the extracted 5D feature space on real fMRI data, *quasi-hemodynamic curve* (QHC) (section 4.2.1) were derived. Although the QHC varies across subjects, brain regions, and cognitive tasks, the characteristics of QHCs often comprise of similar discriminating features. This is illustrated by Fig. 4.8



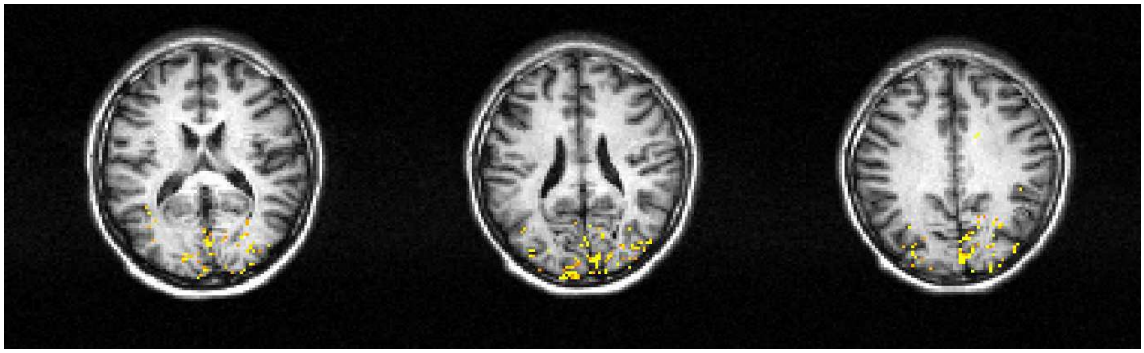
(a) Visual (b) Motor

Figure 4.8: QHC extracted from real fMRI data of (a) visual and (b) motor tasks for both activated (top) and non-activated (bottom) voxels, respectively.

showing typical QHCs for activated (top) and non-activated (bottom) voxels in real fMRI data of (a) visual and (b) motor task, respectively. Despite of the differences between QHCs of the two tasks, evidently, the activated and non-activated voxels have discriminating QHC shapes for both brain regions. Therefore, common essential features are discovered with lower degrees of uncertainty.

Fig. 4.9 shows the detected activated regions for visual task data on three axial slices of one subject by SPM, FFM, cFCM, and wcFCM approaches. Since the ground truth of brain activation is unknown, it is rather difficult to compare activation patterns. Nevertheless, activation was found in expected regions of visual cortex in all approaches. Fewer signal were discovered by FFM as it does not model the contextual information. Both cFCM and wcFCM detected the expected activation by incorporating the contextual information and the results by the latter approach was more

(a) SPM



(b) FFM

(c) cFCM

(d) wcFCM

Figure 4.9: Detected activation on three axial slices from real fMRI data of visual task by SPM, FFM, cFCM, and wcFCM.

focal and contiguous. In addition, SPM and wcFCM detected very similar activation patterns. SPM smoothed the image in order to fulfill its assumptions of statistical analysis while wcFCM was applied on original image without smoothing. Therefore, SPM tends to broaden the activated regions from gray matter into white matter, especially in axial slice three (right).

4.5.2 Counting Stroop Task

Real fMRI data obtained from a counting Stroop task was used to investigate the performance of females with fragile X-syndrome on the cognitive interference processing compared to normal subjects. The cognitive interference occurs when the processing of one stimulus feature impedes the simultaneous processing of a second stimulus attribute [154]. The participants include 14 females with fragile X-syndrome and 14 age-matched healthy control females without the fragile X mutation, ranging in age from 10 to 22 (mean age 15.43). The task consisted of 12 alternating experimental (interference) and controlled (neutral) conditions with the resting condition. For both conditions, subjects were instructed to press the button that corresponded to the number of words that appeared on the screen. During the neutral counting task, the word "fish"was presented 1, 2, 3, or 4 times on the screen (15 trials) and during the interference counting task, the words "one", "two", "three", and "four"were presented 1, 2, 3, or 4 times on the screen (15 trials). Stimuli were presented for 1350ms at a rate of one every 2s (TR) for a total of 180 trials (90 interference, 90 neutral). For more details of the experiment, the reader is referred to Tamm *et al.* [155].

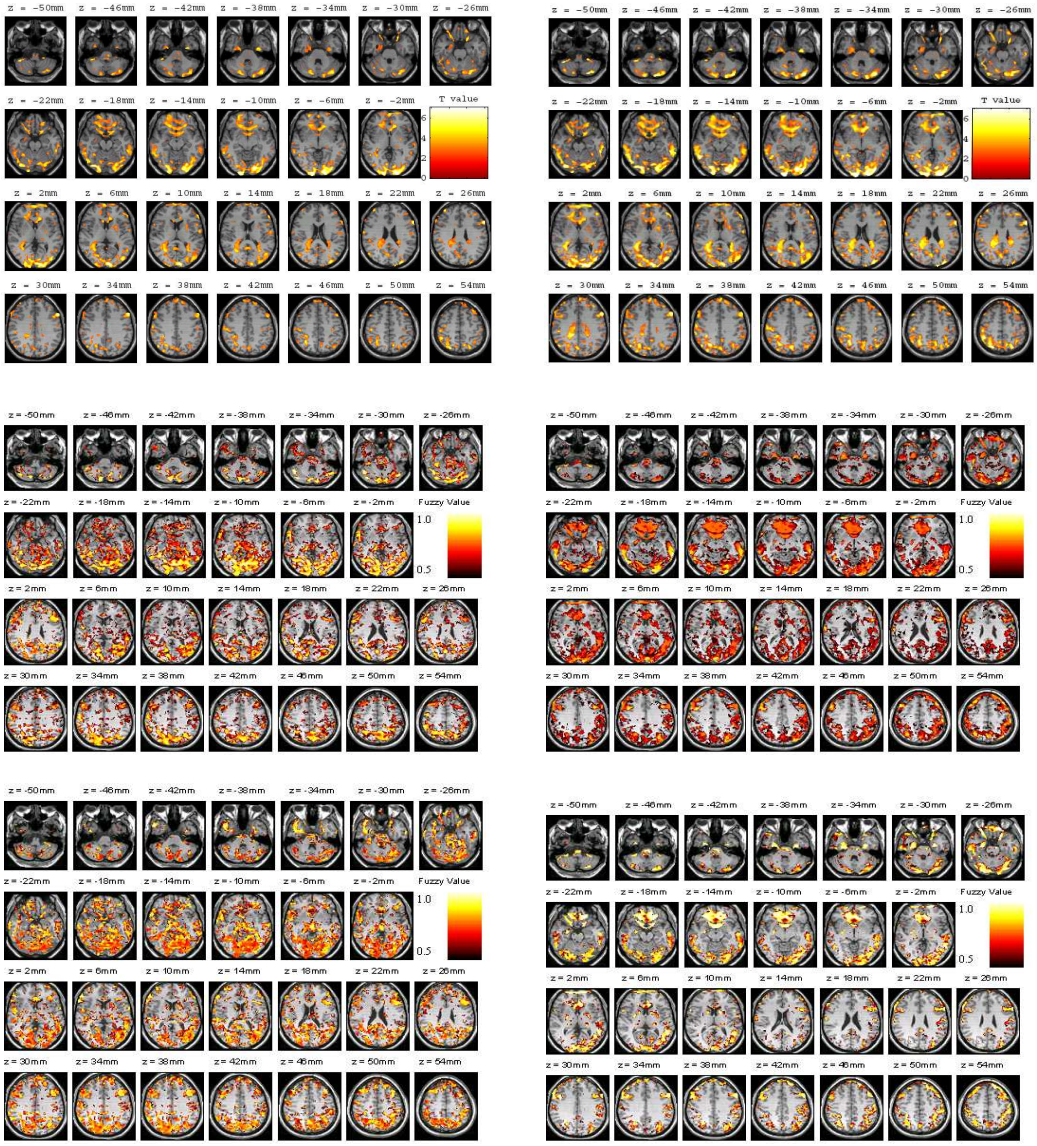
The preprocessed functional images were provided by the fMRIDC; the images were reconstructed by using Inverse Fourier Transform for each of the 225 3D scans into image matrices of $64 \times 64 \times 18$ and voxel size of $3.75 \times 3.75 \times 7$ mm. The regions showing increased activity under neutral and interference conditions were identified by

evaluating the fMRI signal during counting relative to the rest condition. Statistical analysis in SPM was performed based on HRF with time and dispersion derivatives and false discovery rate (FDR) ($p < 0.05$) was used for threshold correction. The proposed FFM and wcFCM approaches and SPM were all applied on the preprocessed images. Parameter settings of FFM and wcFCM were same as those used for synthetic data.

Here, the fMRI data of a female patient was examined to detect the activated brain regions under both neutral and interference counting Stroop conditions. Fig. 4.10 shows the results by SPM, FFM, and wcFCM for each condition, respectively. As observed, their results were very similar. The findings were consistent with previous literature [155] that unique activated brain regions for interference condition include left inferior/middle frontal gyrus (BA 45, 46), left supplementary motor area (BA 6), and right middle/inferior frontal gyrus (BA 9/47). Left and right putamen, left hippocampus, left parahippocampal gyrus, right superior temporal gyrus (BA 22), and right posterior insula were activated in neutral condition.

4.5.3 Group Study of Silent Reading Task

The real fMRI data obtained from a group study of silent reading task were used to evaluate the performance of the proposed approaches on a group of subjects. The dataset consists of six subjects, (five males and one female), aged between 20 and 34, with English as the first language. The experiment consisted of a 3×2 factorial design, three frequencies of presentation: 20, 40, 60 words per minute, and for each, words and pseudowords presentations alternated with a resting condition. The task involved silent reading of words and pseudowords as soon as they appeared on the screen; the resting condition involved fixating to a cross in the middle of the screen. Each subject was presented with 105 words and 105 pseudowords. Stimuli were composed of four,

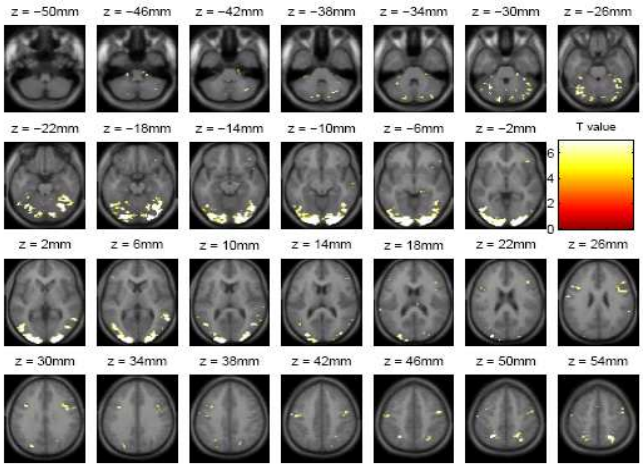


(a) Neutral (b) Interference

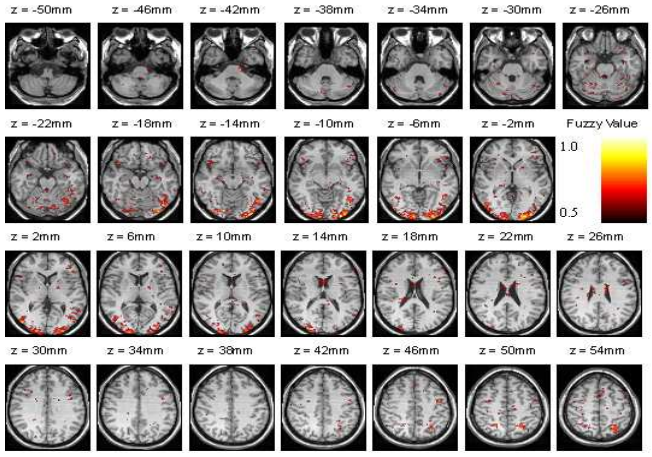
Figure 4.10: Detected activation on selected axial slices from real fMRI data of counting Stroop task by SPM (row 1), FFM (row 2), and wcFCM (row 3): (a) neutral condition and (b) interference condition.

five, or six letters and were presented in 12 blocks. Each block lasted 21s and was followed by a resting period of 16s. The data collected from each subject contains 360 3D scans with a repetition time (TR) of 3.15s. The raw functional MR images were realigned, coregistered, normalized, and smoothed as the preprocessing steps. The time-series were high-pass filtered using a set of discrete cosine basis functions with a cutoff period of 156s and low-pass filtered using a symmetric HRF as the smoothing kernel to condition the temporal autocorrelations (see [156] for details).

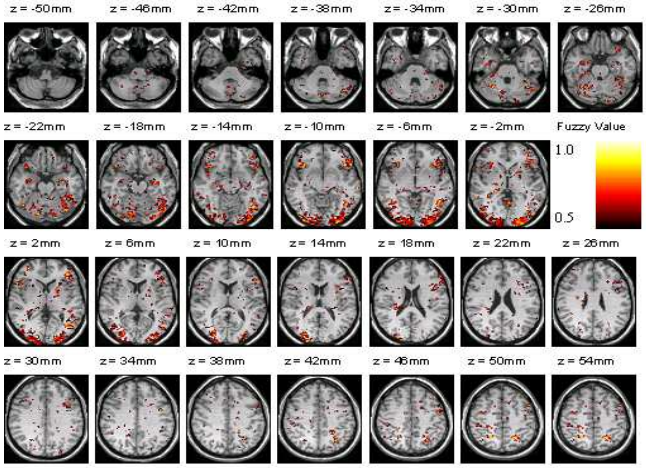
The objective is to identify common activated brain regions during silent reading for both words and pseudowords relative to the rest condition from fMRI time-series of six subjects. The proposed FFM and wcFCM approaches and SPM were all applied on the preprocessed images. In SPM analysis, the design matrix, convolved with the canonical hemodynamic response function (HRF), was used as the reference waveform of each time-series for general linear modelling. Group analyses were performed using a fixed-effect analysis (FFX) [153], which infers typical characteristics about the group of subjects. Significant hemodynamic changes were assessed using the t-statistical parametric maps and statistical inferences were made to report activations at $p < 0.05$ corrected for multiple comparisons by using familywise error rate (FWER) [95, 96]. Parameter settings of the proposed FFM, cFCM, and wcFCM approaches were same as for the synthetic data. The fuzzy activation maps of each subject were combined to determine the common activated brain regions. Fig. 4.11 shows the significant activated brain regions detected by SPM, FFM, and wcFCM. The results were very similar with each other: the activations were found in bilateral extrastraite cortices, superior parietal lobes, middle temporal cortices, inferior frontal sulei, middle frontal cortices and cerebellum, which are consistent with previous literature [15].



(a) SPM



(b) FFM



(c) wcFCM

Figure 4.11: Detected activation on selected axial slices from a group study of silent reading task by (a) SPM, (b) FFM, and (c) wcFCM.

4.6 Discussion

With the novel feature extraction technique, we are able to convert 4D fMRI dataset into a much simpler and robust feature space which is capable of handling the variability of HRF while being less vulnerable to noise and artifacts. Based on the feature space, two fuzzy approaches were developed, fuzzy feature modelling (FFM) and weighted contextual fuzzy c-means clustering (wcFCM), for the detection of functional activation from fMRI data. Fuzzy models were first built on limited prior knowledge represented by the initial set of training data, and iteratively learned to account for different hemodynamic response functions across subjects. A two-way thresholding scheme was proposed to obtain crisp activated regions from fuzzy activation map in an effective manner. In addition to modelling the variability of HRF, unsupervised weighted contextual fuzzy c-means clustering (wcFCM) was developed to incorporate the contextual information of neighbors adaptively. Inter-subject variability of HRF was taken into account by performing data-driven clustering individually on each subject. No thresholding was required as the voxels were automatically assigned to the class (activated or non-activated) with the highest fuzzy membership value.

Synthetic fMRI data was generated to simulate inter-subject variability of HRF and different levels of independent Gaussian and correlated noise. Experiments on synthetic data showed that the proposed wcFCM approach was less vulnerable to both independent and correlated noise as well as inter-subject variability of HRF. It outperformed SPM and FFM and discovered more detailed activation with fewer false positives. The main advantages of our methods were due to modelling of both contextual information and variability of HRF. The proposed FFM approach was less vulnerable to correlated noise but suffered from high level of Gaussian noise because the neighboring voxels are not taken into consideration. Moreover, the resulting centroids of activated and nonactivated classes in FFM and wcFCM could give more

meaningful characteristics and insights of hemodynamic responses. Three sets of real fMRI data obtained from the experiments of visual, counting Stroop, and silent reading tasks were used to evaluate the performance of two proposed approaches by comparing with SPM. It was observed that wcFCM provided more focal and contiguous activated blobs. Both individual and group studies proved that proposed approaches were able to discover brain functional activation from fMRI time-series, which were consistent with previous literature.

Future work could include the accounting of the spatial variability of hemodynamic response function within a single subject. Currently, only two classes (activated and non-activated) are considered for global clustering or learning for each subject though hemodynamic response actually varies across different brain regions. Thus, current approach could be extended to perform global (whole brain) and local clustering (within each brain region) in a hierarchical manner for accurate detection of activation. In this case, the anatomical information such as location, shape, size, and orientation, of brain structures are the key components of such approaches. Multi-modality fusion of structural and functional MR images may be needed [136]. Moreover, tissue information in terms of gray and white matter could be considered as priors for activation detection. Because brain activity are usually located in gray matter tissue rather than white matter or CSF [157], anatomical priors incorporating the variability of brain tissues could potentially increases the accuracy of the present detection approaches.

Chapter 5

Learning Effective Brain Connectivity

Considering the nature of the human brain where multiple processes are taking place and interacting with one another for a specific task, it is of paramount importance to extract brain connectivity from fMRI data [158]. Effective connectivity was recently defined to describe the underlying mechanisms of brain mental processes. In this chapter, we propose to use dynamic Bayesian networks (DBN) to learn the structure of effective brain connectivity from fMRI data in an exploratory manner. The proposed approach does not require any prior model and a MCMC-simulation based structure learning method is used to find an optimal solution in the global sense. Previous exploratory approaches like Bayesian networks (BN) provide a snapshot of effective connectivity of the brain for the entire experiment, and therefore are unable to accurately represent the temporal relationships of connectivity among the regions. Dynamic Bayesian networks (DBN) is capable of capturing temporal relationships among brain regions to infer causal effects. Experiments on synthetic fMRI data demonstrated that the performance of the present method was comparable to Granger causality mapping (GCM) in determining the structure of linearly connected networks. The effective connectivity derived with DBN is more informative as the connections are described in complete statistical sense with the probability density

functions. The functional structures inferred on two real fMRI datasets were consistent with the previous literature and more accurate than those discovered by BN. Furthermore, the robustness of DBN approach was evaluated by studying the effects of hemodynamic noise, scanner noise, and sampling in fMRI experiments. An extension of DBN was also proposed to handle inter-subject variability of effective brain connectivity.

5.1 Introduction

The brain areas involved in various cognitive tasks can now be identified quite accurately and reliably through functional magnetic resonance imaging (fMRI) experiments [88, 106, 107]. However, functional specialization of the brain does not provide the holistic view of brain functions, describing how different brain regions communicate and interact with one another [43]. Considering the nature of human brain where multiple processes take place and interact with one another in executing a specific task, extracting brain connectivity from fMRI could enhance our understanding of brain functions [158]. Recently, there have been an increasing interest in functional integration studies to infer brain connectivity, especially for high-order brain functions [12, 14, 113, 159, 15]. In fMRI, the response of every brain voxel is measured by a time-series of signals based on the blood-oxygenation-level-dependent (BOLD) contrast, due to an increase in blood oxygenation followed by neuronal activity. Given the multivariate voxel-based time series, several techniques have been proposed to use fMRI to characterize the effective connectivity, defined as ``the influence that one neural system exerts over another either directly or indirectly''[42].

Structural equation modelling (SEM) decomposes interregional covariances of fMRI time-series to determine functional interactions among brain regions [13, 113, 121]. The covariance structure models the interactions of the neural system only in the

second-order statistical sense and therefore is unable to render effective connectivity giving the "cause and effect"relationships among brain regions. A combination of ICA and SEM approach was proposed to infer higher-order functional connectivity from fMRI [159]. This is a two-stage approach where ICA is first used to detect and extract task-related activated components and SEM is then applied to infer connectivity among activated brain regions. Dynamical causal modelling (DCM) characterizes the dynamics of interactions among states (brain regions) with bilinear approximations of intrinsic coupling (among neuronal states) and the influence of external inputs (on both states and intrinsic coupling among states). An extended balloon model is used in DCM to model the hemodynamic response, which enables inference of interactions at the neuronal level [12]. Both SEM and DCM are confirmatory in the sense that the analyses of brain connectivity require a prior model to begin with and are not applicable for higher-order functions unique for human such as language or cognition [13].

Granger causality mapping (GCM) extends the vector autoregressive (VAR) technique to capture interactions among brain regions, assuming a causal and dynamic linear system driven by stochastic innovations and linear interactions [14, 160]. Moreover, a graphical approach linking the notions of graphical modelling and Granger causality is applied to describe the dynamic dependencies in neural systems [126]. Nevertheless, a multi-step procedure, fitting autoregression models at each step, is required to identify graphs, limiting its applicability for larger networks or comparison among different networks.

Recently, Bayesian networks (BN) [15] was proposed to derive effective connectivity of the brain from functional MR images in an exploratory manner. However, BN does not provide an explicit mechanism to represent temporal dependencies among multiple processes at brain regions and instead, gives one snapshot of brain connectivity, taking into consideration of the whole experiment. The lack of temporal

precedence information of the connection prevents BN from fully describing ˜cause and effect˝relationships among brain regions. Also, because of equivalent properties of networks, the directions of some edges are indeterminate and could be bi-directional [161].

In this thesis, we proposed dynamic Bayesian networks (DBN) to represent effective connectivity of the brain, that incorporates temporal information of fMRI time-series. DBN, an extension of BN, admits a class of nonlinear continuous-time interactions among states and provides a direct mechanism to model temporal relationships among brain regions. The time-series are modelled with a first-order stationary Markov chain. The inter-scan time difference is used as the difference between two instances of the Markov model. The connectivity between two instances (or scans) of the Markov chain is modelled by a *transition network* with two layers brain regions (or nodes). For example, a connection from node A in layer t to node B in layer representing instant $t+1$ means that node 1 influences node 2 in the next time instance. Two layers of discrete nodes represent observations at two consecutive brain scans and edges represent causal effects. With stationary and first-order assumptions, DBN is able to accurately represent effective brain connectivity by explicitly incorporating temporal relationships.

DBN can assume a known or unknown structure, and full or partial observability of the states of regions. In the case of partially observed states, hidden Markov models are employed [162]. We presume an unknown connectivity structure with full observability and finds the best structure fitting the data in an exploratory manner. The states of brain regions are fully observed as intensity variations of fMR time-series. A greedy search or an expectation maximization (EM) provides only a local search of the structure of DBN. Starting with a fully unconnected structure, we use a Markov chain Monte Carlo (MCMC)-based structure learning method to infer effective connectivity among brain regions from fMRI data of a specific task. A MCMC simulation

attempts to find a globally optimal solution by sampling structures from a Markov chain in equilibrium distribution [163].

We describe the theory behind DBN and structure learning algorithm in the method section, followed by the extension of DBN to model the variability of effective connectivity. In experiments, synthetic fMRI data are used to illustrate the robustness of our approach and compare with earlier technique, GCM. The method is further demonstrated by exploring the functional structure from real fMRI data obtained in two experiments: a silent word reading task and a counting Stroop task; a comparison between structures derived from BN and DBN is provided. The connectivity structure derived by DBN are consistent according to the previous literature.

5.2 Method

This section first describes the application of DBN for studying effective brain connectivity from fMRI time-series. Then a structure learning approach based on MCMC method is described.

5.2.1 Neural System Modelling with DBN

In BN and DBN, we associate nodes with activated brain regions and the observations with average fMRI time-series of the regions while edges characterize interactions among the regions. Considering a neural system R with n brain regions, $R = \{r_i : i = 1, 2, ..n\}$, underlying a cognitive or sensory task. Let fMR time-series x_i, obtained by averaging over time-series of all activated voxels in the brain region, represents the activation of region r_i in terms of the hemodynamic response.

Bayesian networks (BN) describe the probability distribution over a fixed set of vari-

ables and the graphical structure provides an easy way to present and specify conditional interdependencies for a compact parameterization of the distribution. The BN includes a structure S and a joint distribution defined by a set of parameters, θ, over a set of random variables (activations at the nodes), $x = \{x_i : i = 1, 2, ...n\}$. The edges represent conditional dependence relations among the variables. BN is a directed acyclic graph (DAG) characterized by the absence of directed cycles. If a_i is the set of parents of random variable x_i, a DAG offers a simple and unique rule for expanding the joint probability in terms of simpler conditional probabilities:

$$P(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} P(x_i|a_i). \tag{5.1}$$

That is, the network is characterized by a set of conditional probabilities of variables given those of parents.

DBN extends the above representation of the model to incorporate the temporal characteristics of each variable. The set of variables x is characterized by the time variable t to explicitly represent the temporal processes of the brain regions. Let $x_i(t)$ be the random variable representing the activation of region r_i at time t. $x(t) = \{x_i(t) : i = 1, ...n\}$ represents the activations of n brain regions at time t. In order to model the temporal dynamics of brain processes, we need to model a probability distribution over the set of random variables, $\bigcup_{t=0}^{T} x(t)$, which is complex and impractical. The time instances are coupled with scanning times and T represents the total number of scans taken. To avoid an explosion of the model complexity, we assume stationary first-order temporal process, i.e., the transition probabilities $P(x(t+1)|x(t))$ of activations between scans taken at time t and $t+1$ are independent of t.

We assume the temporal changes of activations of brain regions is first-order Markovian, i.e., $P(x(t+1)|x(t), ...x(0)) = P(x(t+1)|x(t))$. In other words, DBN assumes that the connectivity structure between two scans remains same throughout the ex-

periment. Thus, a structure of two layers of nodes representing brain regions in consecutive brain scans is constructed in our DBN, which represents the joint distribution over all possible trajectories of processes. In practice, we only reckon a discrete finite interval $t = [0, T]$, so the structure of DBN can be unrolled into a BN by repeating a transition network for consecutive scans for all t. This unrolling procedure automatically meets the acyclic constraint. The joint distribution over $(x(0), x(1), ...x(T))$ of DBN can be estimated in the same manner as in Eq. 5.1.

The first-order stationary assumption provides a simple causal model that explicitly takes into account temporal dependencies of brain processes. Higher-order and non-stationary Markov models allow more complex temporal processes and connectivity patterns. However, such complex models pose practical challenges in estimating structures and parameters. As seen later, the first-order Markov and stationarity assumptions are valid enough to accurately render brain connectivity in real fMRI experiments.

5.2.2 Structure Learning in DBN

The Bayesian approach to structure learning attempts to find the structure S^* that best fits the available data D; the optimal structure of effective brain connectivity is given by the maximum a posteriori (MAP) estimation:

$$S^* = \arg\max_S P(S|D) \tag{5.2}$$

From Bayes rule,

$$P(S|D) = \frac{1}{Z} P(D|S)P(S) \tag{5.3}$$

where $Z = \sum_S P(D|S)P(S)$ is a normalization factor and $P(S)$ is the prior probability on network structures. The marginal likelihood for a particular network structure,

which requires the parameters θ to be integrated out, is given by:

$$P(D|S) = \int P(D|\theta, S)P(\theta|S)d\theta \tag{5.4}$$

Assumptions

In our structure learning, we make the following assumptions. First, as opposed to Friedman et al. [164], intra-scan connections, i.e., edges within the same brain scan at one time point are not allowed because this would correspond to instantaneous interactions. The interaction in a brain network is not instantaneous as its effect takes place with a time delay after its cause. Although instantaneous interactions may exist in fMRI data due to low temporal sampling, the directions of these interactions are undeterminate and it is still controversial whether the amount of leakage is significant enough to truly reflect the relationship. Moreover, the absence of intra-scan connections enables DBN to model recurrent networks without conflicting with acyclic constraint, which is an important advantage of current DBN model explained in more detail in section 5.2.3. Secondly, inter-scan connections to the same brain region itself are considered as fixed prior connections in the network structure learning process where no deletion is allowed to these self-connections since we assume that the current state of one region automatically influence its state in the future.

Opposed to linear Gaussian distribution, multinomial distribution captures non-linear dependence relations, which is more suitable for modelling effective connectivity in the human brain and is therefore considered in the present algorithm. By this assumption, the integral in Eq. 5.4 is analytically tractable [163] and the Bayesian score is used to measure the fitness of structure. However, this choice requires a discretization of the data and hence suffers from a certain degree of information loss.

MCMC Simulation

The number of network structures increases super-exponentially with the number of nodes and hence the optimization problem is NP-hard. Some heuristic local search algorithms like hill-climbing or simulated annealing has been attempted. Friedman developed the structural EM algorithm for learning both BN and DBN with complete or missing data, which is still a local search algorithm [165, 164].

In contrast, we use a Markov Chain Monte Carlo (MCMC) simulation algorithm for structure learning in DBN in a global search sense [166]. Because of the complex and large model space, $P(S|D)$ may not be adequately characterized by a single structure S. It is more appropriate to sample structures from the posterior probability as in Eq. 5.3, leading to a collection of structures with high posterior probability. This results in structures that offer a good explanation of the data. Since direct sampling from the posterior probability in Eq. 5.3 is impossible due to the intractability of the denominator, we adopt the following MCMC simulation [163]:

Given a network structure S^{old}, a new structure S^{new} is proposed with the proposal probability $Q(S^{new}|S^{old})$. This new structure is accepted with the Metropolis-Hastings (MH) acceptance criterion [167]:

$$min\left\{1, \frac{P(S^{new}|D)}{P(S^{old}|D)} \times \frac{Q(S^{old}|S^{new})}{Q(S^{new}|S^{old})}\right\} \tag{5.5}$$

The iteration of the above procedure generates a Markov chain that under fairly general conditions converges in distribution to the true posterior distribution in Eq. 5.3. In practice, a new network structure is usually proposed by applying one of the elementary operations such as deleting, reversing, or adding an edge and then discarding those structures that violate the acyclicity condition. The Hastings ratio,

in this case, is given by

$$\frac{Q(S^{old}|S^{new})}{Q(S^{new}|S^{old})} = \frac{N(S^{old})}{N(S^{new})} \tag{5.6}$$

where $N(S)$ is the size of the neighborhood of structure S, i.e., the number of acyclic structures that can be obtained from S by application of one of the elementary operations.

The parameters of MCMC simulation are: (1) the length of the burn-in phase, i.e., the number of steps to take before drawing samples, (2) the length of the sampling phase, i.e., the number of samples to draw from the network chain after burn-in phase, (3) the length of the interval between two consecutive samples from the chain after burn-in phase, (4) the maximum fan-in (the number of connections coming into the node) allowed, and (5) the prior expectation of nodes with non-zero fan-out (the number of connections starting from the node). Uniform Dirichlet priors is assumed for parameters of conditional multinomial distributions. The Bayesian score is used to measure the fitness of structure, which integrates over all parameters.

The detailed MCMC algorithm of structure learning is shown as follows:

Algorithm

1. *Initialization*: Initialize the network structure S with all connections set to zero except that intra-scan self-connections were set to 1 as prior knowledge. Discretize the time-series x_i of each region separately into ternary form $[-1, 0, 1]$ by:

$$d_i(t) = \begin{cases} 1 & \text{if } x_i(t) >= \bar{x_i} + (x_{i,max} - \bar{x_i})/3 \\ -1 & \text{if } x_i(t) <= \bar{x_i} - (\bar{x_i} - x_{i,min})/3 \\ 0 & \text{otherwise} \end{cases} \tag{5.7}$$

where $\bar{x_i}$, $x_{i,min}$, $x_{i,max}$ are the mean, minimum, and maximum values of the time-series of region i.

2. *Burn-in phase*: Propose a new network structure by applying elementary move (add, delete or reverse edges) to the old network. Structures violating the acyclic constraint are discarded. Generate a random number u between $[0, 1]$ and the new structure is accepted if u is less than the Metropolis-Hastings (MH) acceptance criterion defined in Eq. 5.5. Repeat the same procedure for a long time to achieve convergence to the true posterior probability.

3. *Sampling phase*: Follow the same procedure as in step 2 and collect the samples for every fixed interval.

4. *Inference*: Calculate the final structure by averaging over all sampled structures in step 3 and the weight on each edge represents its posterior probability.

5.2.3 DBN vs. BN

DBN have two advantages compared to BN for learning effective brain connectivity from fMRI data. Firstly, the intrinsic equivalence property in BN results in loss of information about edge directions because several network structures with the same skeleton but different edge directions can have the same marginal likelihood $P(D|S)$ [161]. Moreover, the lack of temporal precedence prevents BN from inferring important causal interactions among brain regions. In contrast, DBN avoids the ambiguity of the edge directions, i.e., by explicitly taking into account the temporal relationships among brain regions in consecutive brain scans. Since the cause of any event always takes place before its effects, the directions of the edges are always forward in our DBN modelling. For example, if there is a causal relationship from node A to node B, DBN represents this relationship by the existence of an edge $A(t) \rightarrow B(t+1)$ and the absence of an edge $B(t+1) \rightarrow A(t)$ where $A(t)$ and $A(t+1)$ denote the node A in layers representing time instances t and $t+1$, respectively, in the transition network.

Furthermore, the structure in BN must satisfy acyclicity constraint, which rules out recurrent structures, that is, information feedback to itself say after a time interval. As feedback is an essential feature of biological systems, for example, a cortical-subcortical cycle exists in the human brain, this is a limitation of BN for modelling brain connectivity. However, DBN is capable of modelling recurrent connections as it consists of two layers of nodes, each representing same set of brain regions. A recurrent network with three causal relationships ($A \rightarrow B$, $B \rightarrow C$, and $C \rightarrow A$) can be represented by three edges $A(t) \rightarrow B(t+1)$, $B(t) \rightarrow C(t+1)$, and $C(t) \rightarrow A(t+1)$ in our DBN. Similarly, a direct feedback loop with two nodes can be represented by edges $A(t) \rightarrow B(t+1)$ and $B(t) \rightarrow A(t+1)$. Note that in the transition network, self-feedback are not allowed to the same node in the same layer but only through the connections between the two layers.

5.2.4 Modelling Inter-subject Variability

Group studies are usually performed in fMRI experiments because the differences in executing the tasks by different subjects are accounted for valid inference. It is desirable to combine information from multiple subjects to accept or reject certain hypothesis of brain connectivity. However, Goncalves et al. found that the results of the group analysis may not reflect any of the features seen in individual analysis [168]. This is due to various types of variability in brain connectivity as discussed in section 2.4: 1) intra-subject variability, i.e., different runs of the same subject could have different effective connectivity for a particular task; (2) inter-subject variability, i.e., effective connectivity for a same task could vary among individuals; (3) effective connectivity varies for different cognitive tasks; (4) effective connectivity is time-variant. Therefore, it is important to address these variabilities for learning effective connectivity from fMRI time-series of multiple subjects.

The determination of effective brain connectivity with DBN is extended to model inter-subject variability in this section, which has seldom been investigated before. Although a PET study has estimated connectivity by exploiting subject-to-subject variability [169], most of the previous work either dealt with different subject separately, or treated data from different subjects as if they came from the same subject by concatenating them together [13]. This is not desired especially when the temporal relationship among time-series is critical, such as inferring causal and temporal relationship by DBN. Recently, inter-subject variability was addressed by testing directly the differences between models derived from structural equation modelling (SEM) that do and do not allow a particular connectivity parameter to vary among subjects [121]. Another subject-specific approach based on multivariate autoregressive (MAR) model was proposed to estimate the interactions among brain regions for each subject separately and described the variability by the mean and variance of their MAR coefficients. Effective connectivity at the population-level were inferred on the posterior of the MAR coefficients' means in Variational Bayesian (VB) network [170].

Here, we adopt an alternative approach by presuming that fMRI data of each subject is independent. The performance and characteristics of experiments by each subject are dealt with separately. Instead of estimating the effective connectivity parameters of individual subject, the previous DBN model is extended to a probabilistic framework to produce the maximum a posteriori (MAP) solution of the optimal structure considering all subjects together. A structural learning by MCMC simulation for multiple subjects is proposed, i.e., constructing a Markov chain to estimate the aggregate posterior distribution of the structure given multiple subjects, such that the variability among subjects in connectivity structure is implicitly handled and minimized. Therefore, the resulting structure fits to the group data better.

The time-series of each subject is assumed as independent from each other, denoted by D_n. Suppose there is N subjects, the whole set of time-series is represented by

$D = \{D_1, D_2, ...D_N\}$. The optimal MAP structure for all subjects is represented by S^*, so the equation 5.4 in previous section becomes

$$P(D|S) = P(D_1, D_2, ...D_N|S) = \prod_{n=1}^{N} P(D_n|S) \tag{5.8}$$

where $P(D_n|S)$ is the marginal likelihood for data D_n of subject n given network structure S, which can be calculated by integrating parameters θ_n in structure S as

$$P(D_n|S) = \int P(D_n|\theta_n, S)P(\theta_n|S)d\theta_n \tag{5.9}$$

Note that the structure S is same for all subjects but parameters (conditional probability distribution of a node given its parents) θ are different.

The MCMC simulation is similar to the algorithm described in section 5.2.2. In the burning phase (step 2) and sampling phase (step 3), for a random number u generated between $[0, 1]$, the Metropolis-Hastings (MH) acceptance criterion defined in Eq. 5.5 of a new structure is computed for all subject simultaneously. The structure is only accepted if half of subjects accept it. This technique converges to the true posterior probability of the network structure as well as minimizes the differences among subjects. Finally, the network samples collected in the sampling phase are averaged to produce the final structure for group data.

5.3 Experiments and Results

We illustrate our method with experiments on synthetic data as well as on two real fMRI data sets from the fMRI Data Center, Dartmouth College [171]: silent reading task (access number 2-2000-11189) and counting Stroop task (access number: 2-2000-1123B). All simulations were done in MATLAB, using two toolboxes for DBN:

The Bayes Net Toolbox (BNT) [172] and Inferring Dynamic Bayesian Networks with MCMC toolbox [173].

Based on the parameters described in section 5.2.2, two sets of values were tested $[3000, 3000, 5]$ and $[50000, 50000, 100]$ for burn-in phase length, sampling phase length, and sampling interval. The results of the two sets are approximately the same for synthetic data and only the results of $[3000, 3000, 5]$ are reported here. Longer burn-in phase or sampling phase may be more likely to find the global optimal solution at the expense of longer computation time.

5.3.1 Synthetic Data

Synthetic fMRI datasets were generated to test the robustness and stability of the proposed method in detecting the underlying neural system and to compare with the earlier approach (GCM). The effect of temporal sampling, MR scanner noise, and HRF variability on system performance were evaluated.

Data Generation

In order to investigate whether and to what extent DBN is capable of detecting neuronal interactions, we constructed synthetic datasets emulating two effects of real fMRI time-series: hemodynamic modulation and temporal sampling of BOLD response [128]. The steps (see Fig. 5.1(a)) involved in the generation of fMRI data are:

1. *VAR modelling*: generate time-series x of brain regions with a predefined VAR model and additive Gaussian noise. The time interval between consecutive time instances is equal to inter-scan interval (ISI) of 100ms and the length of time-series $T = 3000$ (i.e., 300s). A first-order vector autoregressive (VAR) model

was used to generate the time series:

$$x(t+1) = Cx(t) + u(t) \tag{5.10}$$

where C denotes the linear connectivity matrix and $u(t)$ is the uncorrelated Gaussian innovation with zero mean and covariance matrix Σ with diagonal elements equal to 1 and off-diagonal elements equal to 0.

$$C = \begin{bmatrix} -0.9 & 0 \\ 0.4 & -0.9 \end{bmatrix} \tag{5.11}$$

For example, the connectivity matrix $C = \{c_{ij}\}_{2x2}$ for two regions ($n = 2$) in Fig. 5.1(b) is set as in Eq. 5.11. The generated synthetic data is illustrated in figures 5.1(c) and (d). We set the connectivity matrix C such that (1) there is a directed connection from region 1 to 2 but not in the reverse direction, so DBN is able to discover the right direction of the edge; (2) the spectral power of all the regions are contained in the lower frequency ranges because hemodynamic filtering erases high-frequency influences. As BN has been shown to be superior than SEM in obtaining the connectivity structure [15], we chose the above model of synthetic data to enable the comparison with GCM.

2. *HRF modulation*: modulate time-series by convolving with the canonical hemodynamic response function (HRF) consisting of a mixture of two gamma functions:

$$f(t) = \frac{1}{\Gamma(\frac{\tau_1}{\tau_3})}(\frac{\delta}{\tau_3})^{\frac{\tau_1}{\tau_3}} w(t)^{\frac{\tau_1}{\tau_3}-1} e^{-\frac{\delta}{\tau_3}w(t)} - \frac{1}{\tau_5\Gamma(\frac{\tau_2}{\tau_4})}(\frac{\delta}{\tau_4})^{\frac{\tau_2}{\tau_4}} w(t)^{\frac{\tau_2}{\tau_4}-1} e^{-\frac{\delta}{\tau_4}w(t)} \tag{5.12}$$

where $\delta = \frac{RT}{16}$, repetition time $RT = 100$ms, delay of response $\tau_1 = 6$s, delay of undershoot $\tau_2 = 16$s, dispersion of response $\tau_3 = 1$s, dispersion of undershoot $\tau_4 = 1$s, ratio of response to undershoot $\tau_5 = 6$, and the length of kernel $\tau_6 = 32$s.

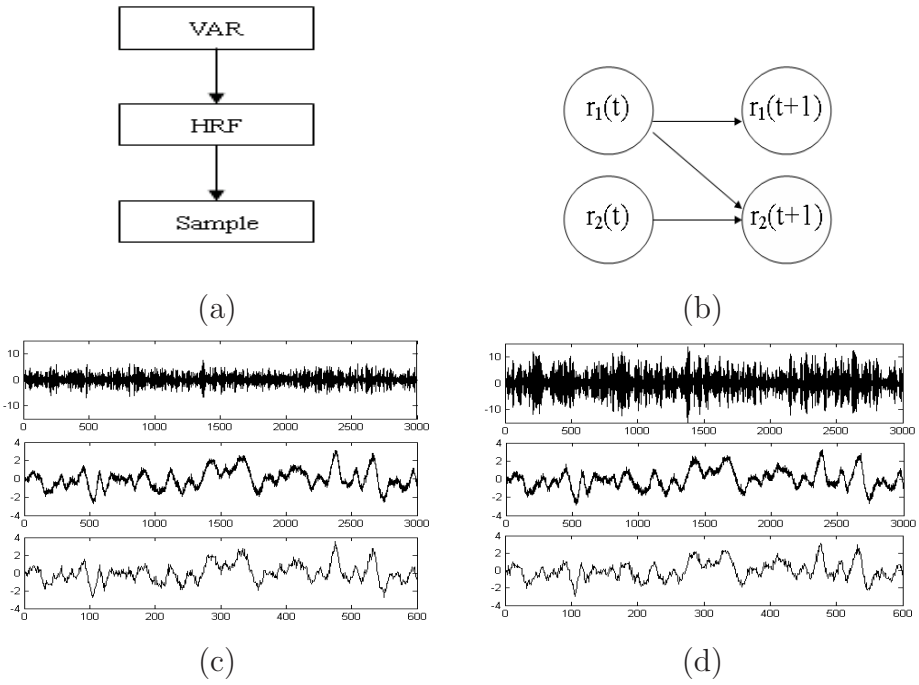


Figure 5.1: Generation of synthetic data: (a) three steps involved in generation of time-series; (b) an example structure of two regions; (c) and (d) correspond to time-series generated at regions r_1 and r_2, respectively.

$\Gamma(\cdot)$ represents the gamma function and $w(t) \in \{0, \frac{1}{\delta}, \ldots \frac{\tau_6}{\delta}\}$. After normalized to zero mean and unit variance, a certain amount σ_1 of white Gaussian noise was added to emulate hemodynamic noise.

3. *MR sampling*: acquire final time-series by down-sampling the signals at every Δ second, corresponding to the ISI of MR scanner. After normalization of signals individually, amount σ_2 of Gaussian noise was added to represent errors in measurement and acquisition (or scanner noise).

Using the above procedure, we generated neural systems consisting of different number of brain regions. The default parameters were $u(t) \sim N(0, 1)$, hemodynamic noise $\sigma_1 = 0$, scanner noise $\sigma_2 = 0$, and inter-scan interval $\Delta = 0.5s$.

Parameter Settings

For MCMC simulation, two sets of parameters were tested: $[3000, 3000, 5]$ and $[50000, 50000, 100]$ for lengths of burn-in phase, sampling phase, and sampling interval, respectively. The networks derived for synthetic data with two sets were approximately the same and only the results of $[3000, 3000, 5]$ are reported here. Longer burn-in phase or sampling phase may be more likely to find the globally optimal solution at the expense of computation time. There is no constraint on the maximum fan-in and a flat prior expectation of nodes with non-zero fan-out is used so that the learned connectivity structure is completely data-driven without any constraints or bias from prior knowledge.

The robustness of our method of deriving DBN for inferring connectivity structure was evaluated with respect to four aspects: (1) the number of brain regions n, (2) the ISI Δ in sampling, (3) the hemodynamic noise σ_1, and (4) the scanner noise σ_2. For different values of above parameters, synthetic fMRI datasets were generated: n from 3 to 8; Δ from 0.5s to 3.0s; the length of time-series T from 600 to 100 points covering 300s and representing fMRI experiments with RT from 0.5s to 3.0s; the mean of both signal and noise were set to zero; the standard deviation of signal was set to 1.00; the standard deviations of noises $\sigma_1, \sigma_2 \in \{0.00, 0.05, \ldots 0.50\}$.

An example network of five regions $n = 5$ is shown in Fig. 5.2. Time-series data of $T = 600$ time points were generated. The execution time in deriving the DBN with MCMC parameters $[3000, 3000, 5]$ was approximately 4min on a Pentium IV 2.4GHz machine having 512MB memory and running MATLAB version 7.0 on Windows XP environment. The GCM approach took approximately 2min under the same condition.

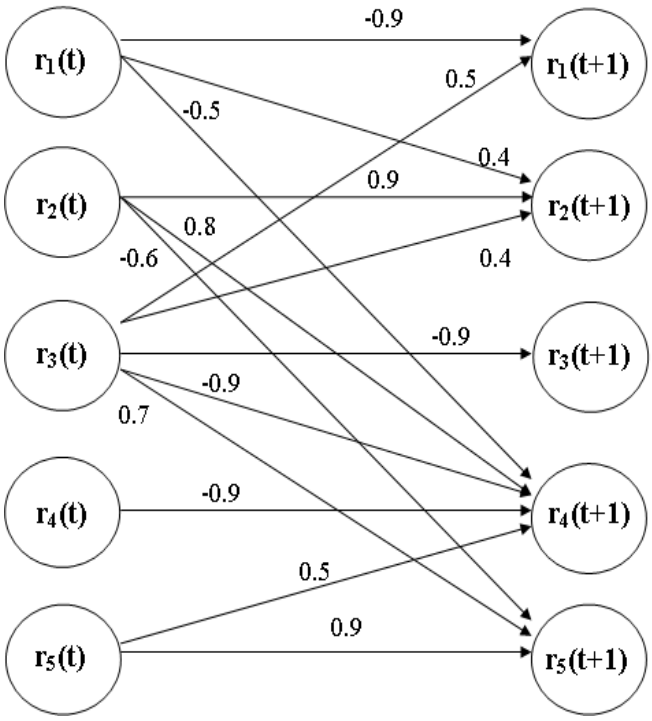


Figure 5.2: Example of a typical brain connectivity structure consisting of five regions. The numbers at the edges represent the strength of connections.

Performance Evaluation

The square error e^2 between the true connectivity structure $C = \{c_{ij}\}$ and the estimated structure $\hat{C} = \{\hat{c}_{ij}\}$ is measured by the square error between the elements of the matrices defining the structures:

$$e^2 = \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (c_{ij} - \hat{c}_{ij})^2; \tag{5.13}$$

the elements C and $\hat{C}$ were scaled to the range $[0, 1]$ for comparison with GCM results. For a given structure s, five different datasets were generated following the above procedure with random initializations. MCMC simulation used the same set of parameters $[3000, 3000, 5]$. The mean and standard deviation of square errors e^2 are reported.

The performance of DBN with various sampling step is shown in Fig. 5.3(a). Each curve in the figure shows the mean of average error $\bar{\varepsilon}$ versus the sampling step Δ on learning the structure. As seen, DBN with MCMC-simulation is capable of deriving neural system from synthetic fMRI data closer to the ground truth with small average error (< 0.1) for all values of sampling step up to 3s. As expected, when the number of regions n increases, the average error increases, which could mean that the algorithm has a higher probability to fall into local minimum. Note that on the whole, sampling step does affect the performance of DBN, which is nevertheless smaller than the effect of the number of regions.

Fig. 5.3(c) and (e) show the mean of average error as the amount of noise, σ_1 of HRF modulation, and σ_2 in MR scanner, increase based on the baseline model. Each curve in the two figures shows the average error $\bar{\varepsilon}$ versus the noise, σ_1 and σ_2, on learning the structure for a dataset with specified number of regions. Effects of both HRF noise and MR scanner noise are more severe than those contributed by sampling and they take the leading role of controlling the average error compared to the number of regions. The standard deviation of the average error $SD(\varepsilon)$ was smaller than 0.025 for all simulations shown in the three figures 5.3(a), (c), and (e), indicating the stability of DBN in estimating the structure of effective connectivity. As expected, when the sampling step, HRF noise, and MR scanner noise increase, the standard deviation of average error increases.

Comparison with GCM

GCM technique [14, 128] was used to derive the neural systems from the synthetic datasets in order to quantitatively compare the performance with the present technique using DBN. We used the algorithm proposed in [174] to estimate parameters in VAR model and Schwarz-Bayesian criterion (SBC) was used to estimate the optimal

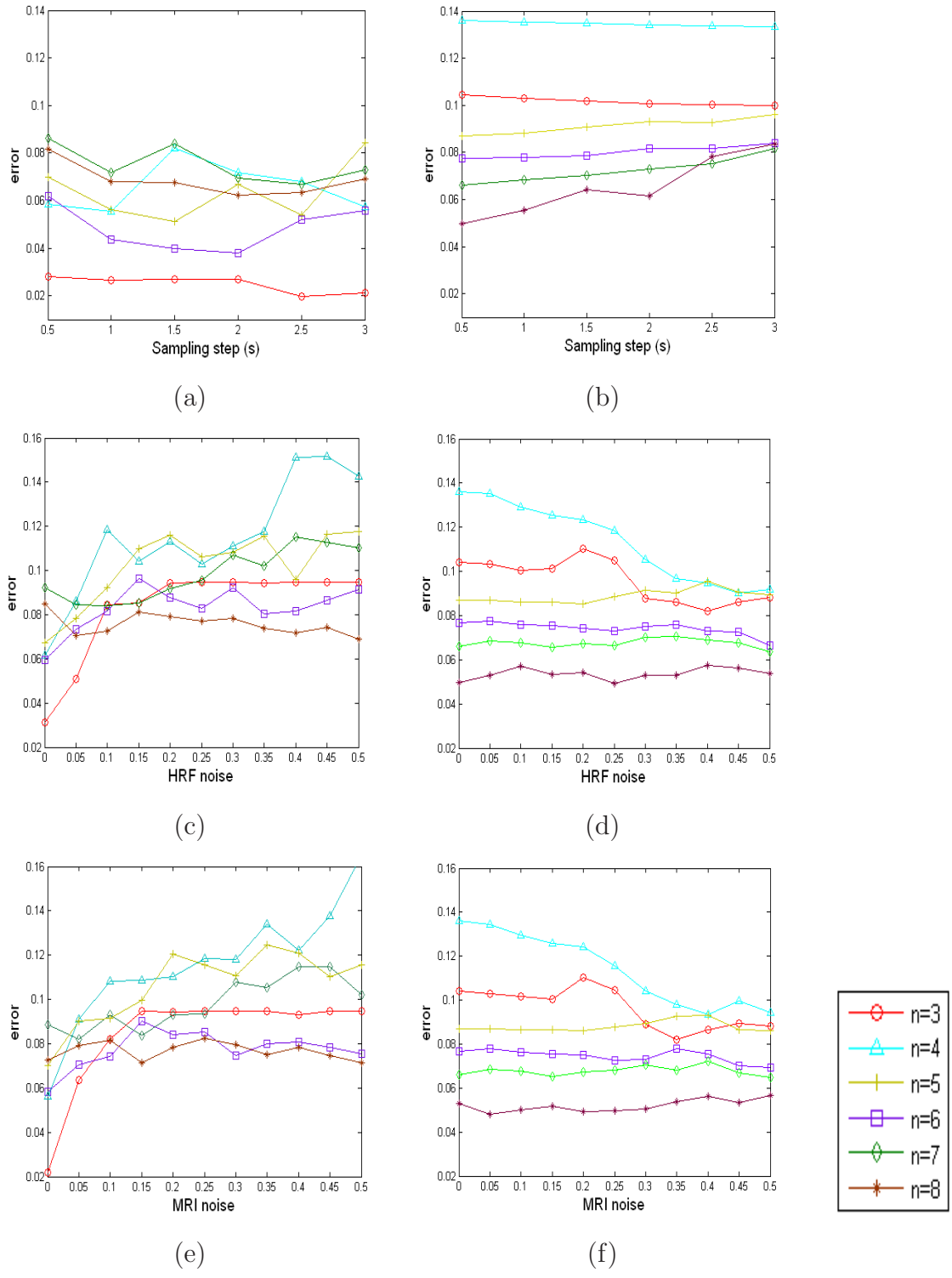


Figure 5.3: Comparing the performance of DBN (left column) and GCM (right column) for structure learning of synthetic data with respect to (a) and (b) sampling steps; (c) and (d) the amount of noise in HRF modulation; and (e) and (f) the amount of noise in MRI scanner sampling.

order of VAR model.

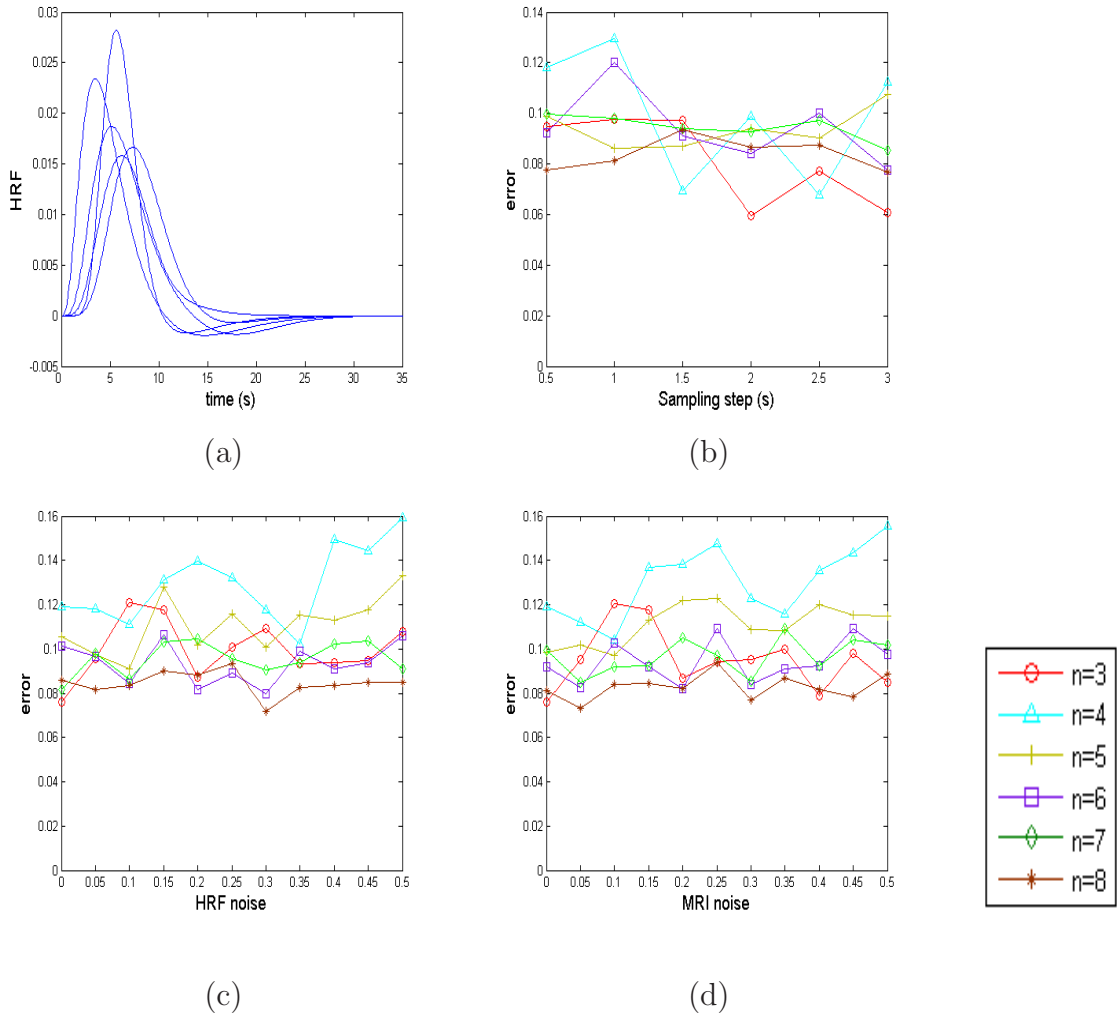
The three figures in the right column, (b), (d), and (f), of Fig. 5.3 illustrate the effects of sampling step, HRF noise, and MRI noise on the performance of the GCM technique, respectively. The order of 1 was always selected as the optimal order based on SBC with GCM. In order to compare properly, we scaled elements of C and $\hat{C}$ to the range $[0, 1]$ in order to compute the average error. Despite of the information loss due to discretization of fMRI time-series, DBN still achieves performances comparable to GCM technique. Both methods had similar effects due to sampling step, i.e., the average error increased with the sampling step, but DBN always showed better performance than GCM at the same sampling rate. Both techniques showed increased error with HRF noise and MRI noise when the number of regions ranged from 3 to 8. GCM is less vulnerable to HRF and MRI noise than DBN especially for the network structure with 3 or 4 regions partly because of the effects of discretization of time-series in DBN. As the noise increases, the loss of information resulting from discretization becomes prominent, influencing the performance of our technique. Interestingly, sometimes larger number of regions showed lower average error for both DBN and GCM. This may indicate that the network structure with a larger number of nodes is less vulnerable to noise and could be better estimated with both GCM and DBN.

Robustness to HRF Variability

The characteristics of HRF vary across brain regions and subjects [25]. In order to evaluate the robustness of DBN to the variation of HRF, different HRF were generated by randomly selecting the values of six parameters, as defined in Eq. 5.12, within their practical ranges: delay of response $\tau_1 \in [4.0, 8.0]$s, delay of undershoot $\tau_2 \in [12.0, 20.0]$s, dispersion of response $\tau_3 \in [0.5, 1.5]$s, dispersion of undershoot

$\tau_4 \in [0.5, 1.5]$s, ratio of response to undershoot $\tau_5 \in [4.0, 8.0]$, and the length of the kernel $\tau_6 \in [28.0, 36.0]$s. The neural activity of every region of different subjects was convolved with a randomly selected HRF. Fig. 5.4 (a) shows five different HRFs generated for a five-regions model in Fig. 5.2.

For the same structure s, at a given sampling step size, hemodynamic noise level, and scanner noise level, five different datasets were generated with random initialization, hemodynamic modulation, and temporal sampling. The synthetic datasets were generated and the effects of sampling, hemodynamic noise, and scanner noise were tested, in the manner as described in section 5.3.1. In order to evaluate the capability of our approach handling the variability of HRF among different brain regions and subjects, we compared the averages of square error in learning structures on data generated with the canonical HRF and with varying HRF parameters. Paired T-test was used to examine whether there is a significant increase in the error for structures learned. Table 5.1 gives p-values of comparisons. As seen, the variability of HRF significantly affects the learning of structures with varying sampling rate but less with hemodynamic or scanner noise. Figures 5.4 (b), (c) and (d) show the average error with respect to sampling step, HRF noise and MRI noise. By comparing to the results in Fig. 5.3 for data without HRF variability, it is observed that HRF variability among voxels affects system performance with respect to sampling step. Convolution with HRF of different delay results in high vulnerability to sampling. However, as the sampling step increases, the error tends to decreases slightly as the delay in HRF is no longer significant compared to large sampling step. For noise in HRF and MRI scanner, we can see that HRF variability influences more on data with lower level of noise. The error for data with higher level of noise remains same as data without HRF variability.



(a) (b)

(c) (d)

Figure 5.4: Robustness of DBN for structure learning of synthetic data based on various HRFs: (a) different HRFs; the error with respect to (b) sampling steps; (c) the amount of noise in HRF modulation; and (d) the amount of noise in MRI scanner sampling.

Regions	Significance of error difference (p-value)		
n	sampling interval	hemodynamic noise	scanner noise
3	$p < 0.01$	$p < 0.01$	–
4	$p = 0.03$	$p = 0.02$	$p = 0.03$
5	$p < 0.01$	$p = 0.04$	–
6	$p < 0.01$	$p = 0.02$	$p < 0.01$
7	$p < 0.01$	–	–
8	$p = 0.01$	$p < 0.01$	$p < 0.01$

Table 5.1: Comparison of error in structure learning on synthetic data, considering HRF variability. A '-' indicates no significant increase in the error of detecting connectivity due to the variability of HRF parameters at the level of $p < 0.05$.

Modelling Inter-subject Variability

In order to test the algorithm described in section 5.2.4, instead of applying DBN to each time-series separately, an optimal structure was found by considering multiple times-series generated as in section 5.3.1, simultaneously. Five sets of synthetic data were generated based on a single true structure with spatial and inter-subject variability of HRF as described in section 5.3.1. The same set of parameters of MCMC simulation was used as in the previous section.

As in section 5.3.1, the extended version of DBN for multi-subject learning was evaluated and figures 5.5 (a), (b), and (c) show the average error with respect to sampling step, HRF noise, and MRI noise, respectively. By comparing current results with the results in Fig. 5.4 of learning effective connectivity from individual subjects, it is observed that the proposed DBN framework is able to discover the true structure from a group of subjects at least with comparable performance to when treating each subject individually. Note that the same length of sampling steps were used for comparison. Moreover, the current approach produces a single common network structure for multiple subjects with minimal variability while the previous approach results in different network structure for each individual.

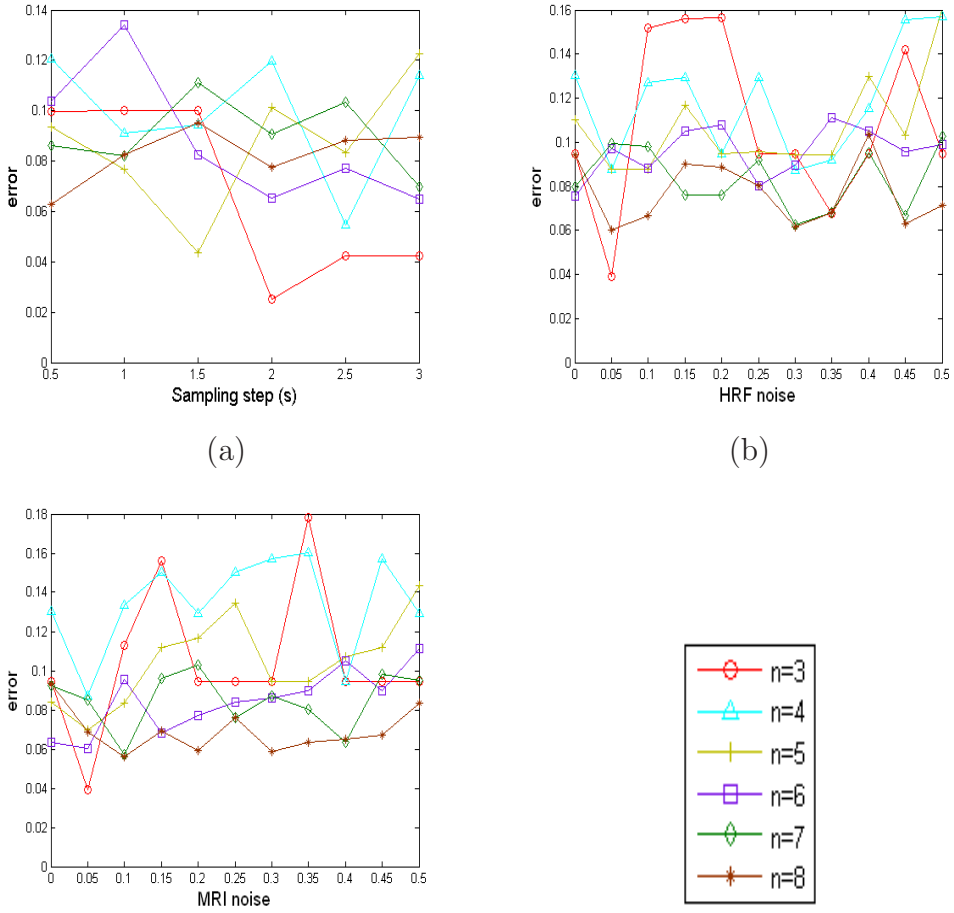


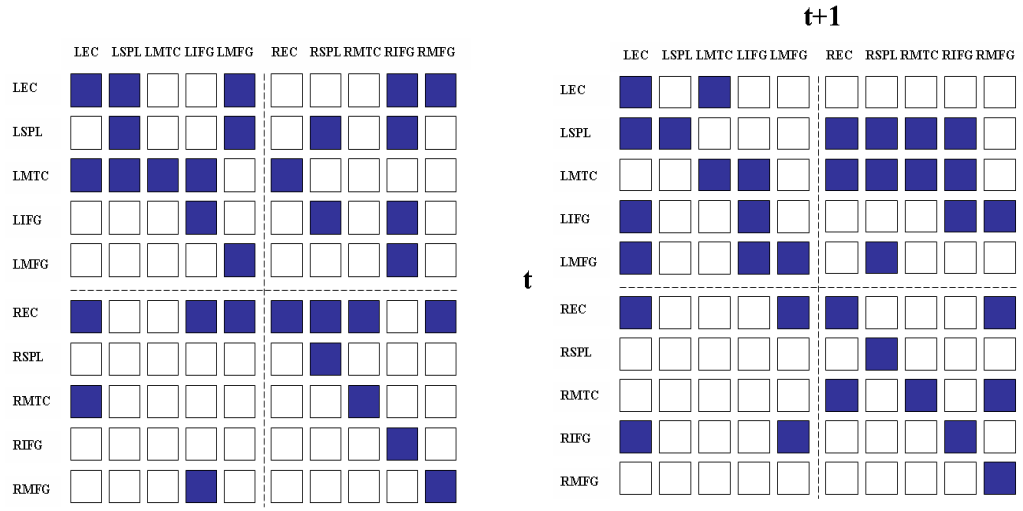
Figure 5.5: Performance of DBN associated approach for structure learning of synthetic data of multiple subjects: the error with respect to (a) sampling steps; (b) the amount of noise in HRF modulation; and (c) the amount of noise in MRI scanner sampling.

5.3.2 Silent Reading Task

FMRI data on a silent reading task were collected from six normal subjects. The task involved silent reading of words and pseudowords as soon as they appeared on the screen. The resting condition involved fixating to a cross in the middle of the screen. The data of each subject contains 360 3D brain images. For more details, please refer to [156].

In order to compare with BN, SPM analysis was performed for activation detection and time-series extraction as described in [15]. The time course of significantly activated regions were extracted by taking the averages of the time-series at the peak-activated voxels and its neighbors at the cluster level for all subjects. The time-series of all subject were concatenated together to form the dataset for deriving the brain connectivity with BN and DBN. Same MCMC simulation algorithm was applied on discretized data. As the number of regions involved in the dataset is large ($n = 10$), we use longer burn-in and sampling phases and a large sample size: $[50000, 50000, 100]$ to avoid trapping into local optima as much as possible.

Fig. 5.6 shows the learned network structures from BN [15] and DBN. DBN selected those edges with the resulting probability greater than 0.5. As seen, similar connectivity structures were found in both BN and DBN. There are more intra-hemisphere connections in left brain than right brain. Note that the diagonal edges (from top left to bottom right) are not compared since each region is assumed to be related to its own history in DBN framework, which is not the case in BN. As explained in section 5.2.3, the directions of edges in BN are undetermined, i.e., a connection from (LEC→LSPL) in the learned structure by BN in Fig. 5.6(a) may not represent the right direction of the cause and effect relationship between the two regions. In contrast, DBN provides edge information correctly by taking into consideration the cause and effect between scans. As seen from the structure derived from DBN, left



(a) BN (b) DBN

Figure 5.6: The neural system learned from fMRI data of silent reading task with (a) BN and (b) DBN. L(R)EC: left(right) extrastriate cortex (BA 18/19), L(R)SPL: left(right) superior parietal lobe (BA 7), L(R)MTC: left(right) middle temporal cortex (BA 21/22), L(R)IFG: left(right) inferior frontal gyrus (BA 44/45), L(R)MFG: left(right) middle frontal gyrus (BA 46/9). A shaded square in the transition diagram indicates the existence of a connection from a region in the row to a region in the column. Rows and columns in DBN represent the brain regions in time instances t and $t+1$, respectively.

hemisphere has more directed connections to the right hemisphere, which is consistent with the previous findings that left hemisphere is more involved in language processing. This conclusion is difficult to be inferred from BN [15] where the direction of the edges of BN could represent bi-directional dependencies.

DBN found several connections previously reported in the literature for language processing: LSPL → RSPL [119] and REC → LEC for homologous interconnection [114], LSPL → RIFG for phonemic decisions [119], LMTC → LIFG for semantic phonologic retrieval [116, 175], LMTC → REC for memory retention [114], REC → LMFG and REC → RMFG for semantic decision and analysis [115]. The temporal relationship of (REC, RMTC) and (LEC, LMTC) was not found in BN most likely due to bi-directional possibility of the connections found [15]. Furthermore, DBN found a recurrent network (LEC → LMTC → RMTC → REC → LEC) associated

with retaining and recalling words from the memory, in which MTC is in charge of integrating the input from the lower level auditory and visual areas for retaining in the memory.

As the extrastriate cortex (EC) and superior parietal lobe (SPL) are important in visual representation and analysis, respectively, in word processing, the functional connection between EC and superior parietal lobe (LSPL → LEC, LSPL → REC) and other connections with prefrontal cortex (LSPL → RIFG, LMFG → RSPL, REC → RMFG, REC → LMFG, RIFG → LEC, LIFG → LEC, LMFG → LEC) enable the perception of visual word form [176]. Connections between MTC and SPL indicate their close relationship in dealing with visual words (LSPL → RMTC, LMTC → RSPL). Within prefrontal cortex, intra- and inter- hemispheric connections were found similar to BN (LIFG → RMFG, LIFG → RIFG). The connection (LMFG → LIFG) found in DBN was absent from BN, which has been earlier reported in an experiment demanding semantic categorization and subvocal rehearsal [13].

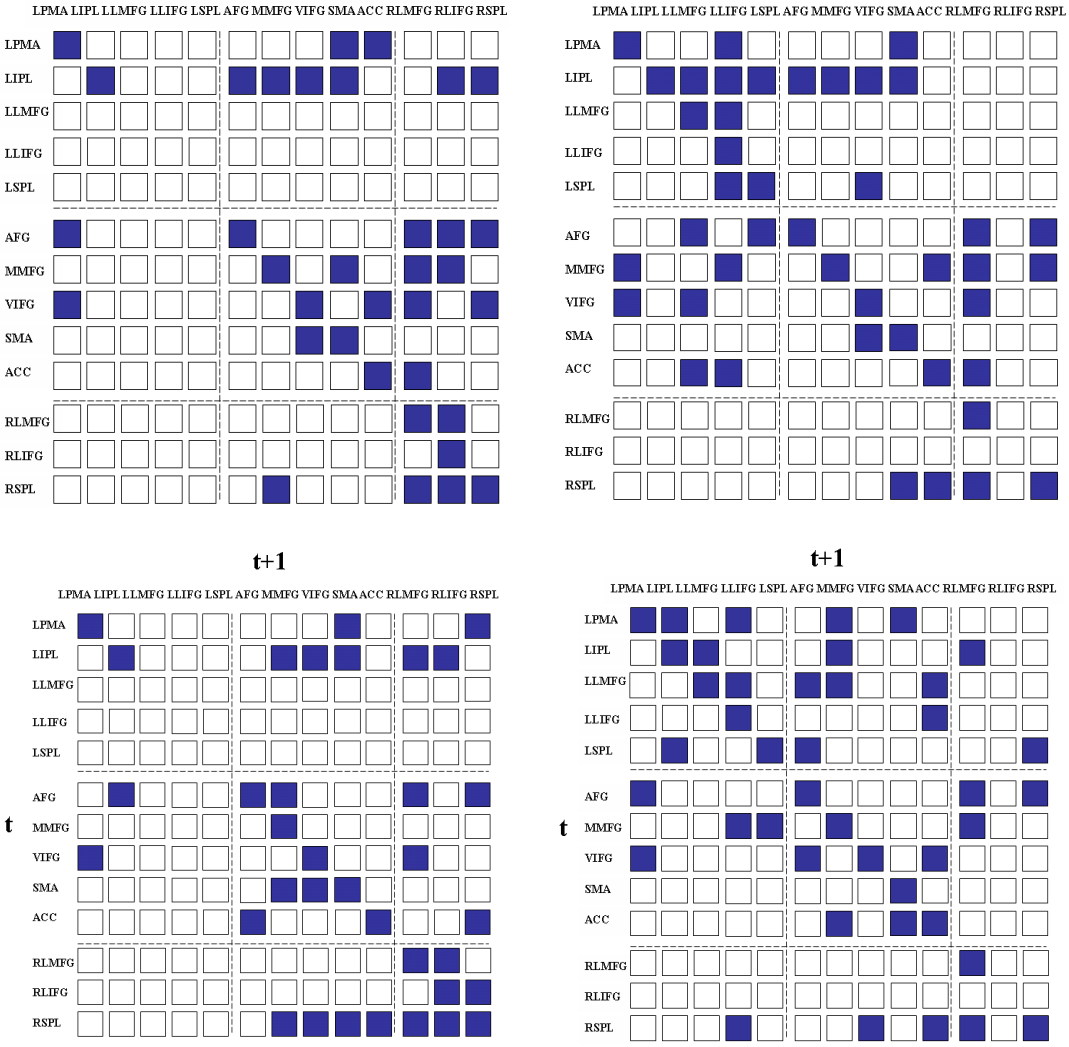
5.3.3 Counting Stroop Task

FMRI data in a counting Stroop task investigating the performance of females with fragile X-syndrome on the cognitive interference processing compared to normal subjects were analyzed. There are two conditions in the task: neutral and interference, see [155] for more details about the task and data collection. Using SPM2, time-series were extracted to derive connectivity using BN and DBN. We explored the networks involved in neutral and interference counting Stroop condition of normal controls and attempted to make a comparison. Ten regions were found activated in neutral condition while twelve regions were found activated in interference conditions. Time-series of 14 normal control subjects were concatenated together to form training data; same settings for MCMC-simulation for learning structure in DBN were used as for silent

reading task.

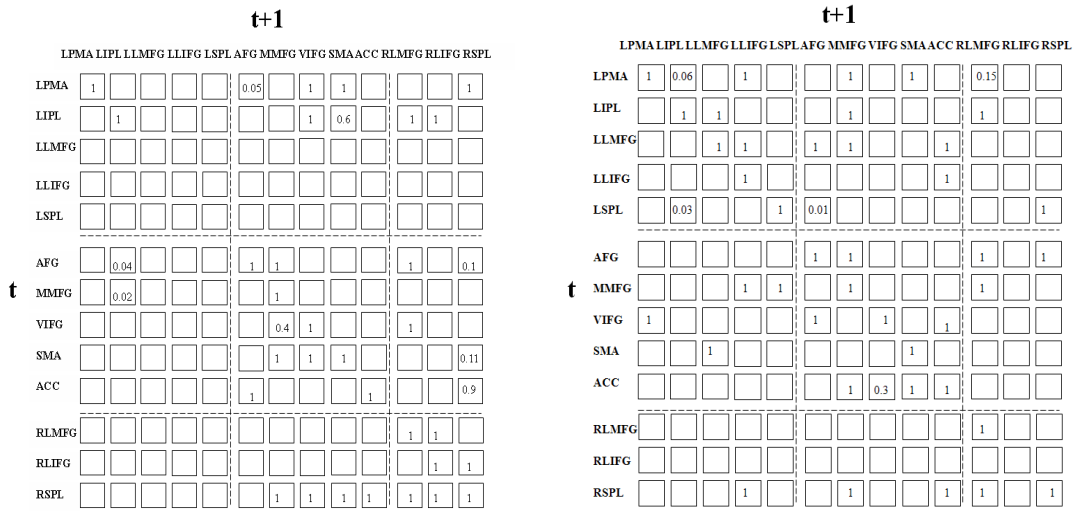
Fig. 5.7 shows the learned structures for neutral and interference conditions by BN (top) [15] and DBN (bottom). Note that the region with the empty diagonal square (from top left to bottom right) means that the region is non-activated in this condition. DBN used the threshold 0.5 for edge connections to determine the crisp network structure. Fig. 5.8 shows the original unthresholded network with its connection strengths derived from DBN. The connectivity structures found by BN and DBN are quite similar: (1) large amount of connections within medial cortices for both conditions, indicating the responsibility of counting function; (2) the number of intra-hemispheric connections in left brain and between left brain and medial cortices was much fewer in neutral condition than interference condition, both indicating the more involvement of language processing and decision making in left hemisphere under interference effects.

Anterior cingulate cortex (ACC) connected with the prefrontal cortex and parietal cortex as well as the motor system and the frontal eye fields is a central station for processing top-down and bottom-up stimuli and assigning appropriate control to other areas in the brain [177]. ACC particularly plays an essential role in counting Stroop task [154, 178, 179] to resolve competing streams of information in the selection of sensory inputs and responses. This is evidenced by connections in the interference network while absent from neutral network: ACC → MMFG, ACC → SMA, LLIFG → ACC, LLIMFG → ACC, VIFG → ACC. Although these connections exist in both BN and DBN, bi-directional characteristics of BN is unable to confirm their actual directions without looking at several optimal structures, starting from different initial conditions. Because more concentration is required for interference task, the connections from ACC to MMFG, which controls the eye movement, are also absent from neutral task. ACC receives the processed input from frontal gyrus (VIFG, LLIMFG, LLIFG) and assigns appropriate motor control to Supplementary Motor Area (SMA).



(a) Neutral (b) Interference

Figure 5.7: The neural system learned from fMRI data of counting Stroop task (neutral and interference condition) by BN (top) and DBN (bottom). LPMA: left primary motor area (BA 4); LIPL: left inferior parietal lobe (BA 40); L(R)LMFG: left (right) lateral middle frontal gyrus (BA 9); L(R)LIFG: left (right) lateral inferior frontal gyrus (BA 44); L(R)SPL: left (right) superior parietal lobe (BA 7); AFG: anterior frontal gyrus (BA 10); MMFG: medial middle frontal gyrus (BA 8); VIFG: ventral inferior frontal gyrus (BA 47); SMA: supplementary motor area (BA 6); ACC: anterior cingulate cortex (BA 24). A shaded square indicates the existence of a connection from the region in row to the region in column. The rows and columns in DBN represent the regions in time t and $t+1$, respectively.



(a) Neutral (b) Interference

Figure 5.8: The unthresholded neural system learned from fMRI data of counting Stroop task (neutral and interference condition) by DBN with their connection strengthes. The rows and columns in DBN represent the regions in time t and $t+1$, respectively. Same abbreviations of brain region are used as in Fig. 5.7.

DBN also discovered the connection RSPL → ACC for both neutral and interference condition. This is consistent with earlier findings about the connection between ACC and parietal cortex [177]. However, BN could not find this connection for interference condition.

Supplementary motor area plays an important role in planning complex and coordinated movements [176] and primary motor area (PMA) stores motor patterns and responds to voluntary activities [180]. VIFG controls the voluntary goal directed behavior [155] and a connection chain (VIFG → LPMA → SMA) for voluntary movement planning is found in both conditions by both BN and DBN. AFG is participated in memory retrieval and executive function and LMFG is responsible for the selection of behavior based on the short term memory, in this case, solving conflict, and LIFG is in charge of phonemic decisions [181]. Both BN and DBN have found similar connectivity structures: AFG → RLMFG exists for both neutral and interference conditions, RLMFG → RLIFG exists only in neutral condition while LLMFG → LLIFG is present only in interference condition. MMFG is believed to control eye movements [180] and

related to uncertainty management [182]. Thus, the connections related to MMFG express different patterns of connectivity under conflicting situations. In neutral condition, the connection SMA → MMFG exists while more involvement of MMFG is found in interference condition such as LPMA → MMFG, LLMFG → MMFG → LLIFG, and MMFG → RLMFG.

The above areas in frontal lobe receive inputs from parietal lobe which generally performs the functions of processing and discriminating of sensory inputs. However, the connectivity structures between frontal gyrus and parietal lobe are different for neutral and interference conditions. Both BN and DBN have found common connections AFG → RSPL → RLMFG and LIPL → MMFG in both neutral and interference condition. In addition, DBN found new connections between frontal gyrus and parietal lobe common to both interference and neutral conditions: LIPL → RLMFG and RSPL → VIFG. Besides, the structure derived by DBN shows an important and specific language pathway (LSPL → LIPL → LLMFG → LLIFG) in the interference counting task for phonetic and semantic analysis and decision. The connections RSPL → LLIFG and MMFG → LSPL are also absent from neutral conditions. Meanwhile, RSPL → RLIFG, LIPL → RLIFG, and RSPL → MMFG only exist in neutral network, which perform a compensatory function for the absent language pathway to visualize the symbols instead of reading, or "automatic speech"in right hemisphere [183].

5.4 Discussion

We used DBN to derive effective connectivity among brain regions from fMRI data. A structure learning algorithm based on MCMC simulation was introduced to search for the best connectivity structure predicted by the data. The validity of present approach was illustrated with experiments on synthetic data emulating true fMRI

time-series characteristics. The performance was comparable with GCM in a linearly connected network. The discovered effective connectivity models from real fMRI data by DBN were consistent with the previous literature and similar to those found by BN. Moreover, the current approach has been extended to model inter-subject variability to learn an optimal structure from a group of subjects together. Experiments on a group of synthetic fMRI data showed comparable performance with learning from the data of each subject separately.

Confirmatory techniques for effective connectivity analysis like DCM are only useful when a prior model is available and thus they are not suitable for complex cognitive tasks or functional integration studies for which prior connectivity patterns are unavailable. GCM only models linear stochastic process and SEM extracts interactions by second-order statistics. Bayesian networks (BN) is in contrast able to explore the optimal connectivity structure from fMRI data automatically without any prior knowledge of the structure. Moreover, BN is able to characterize the effective connectivity among brain regions in a complete statistical sense compared to linear connectivity in VAR models of connectivity assumed by GCM or second order connectivity by SEM. However, the directions of some connections could not be determined because there could be several BNs satisfying the optimal criterion with different edge directions.

Dynamic Bayesian networks (DBN) is capable of learning the structure together with their directional characteristics. Further, BN does not capture the complete picture of the temporal characteristics of fMRI time-series as it collapses all time points into one snapshot depicting the brain connectivity. On the other hand, DBN considers the temporal changes between scans in deriving the connectivity structure. Therefore, DBN offers cause and effect relations among brain regions with more accurate temporal relationships. The advantage of interpreting the temporal and causal relationship more accurately among different processes at brain regions was seen by experiments on real fMRI data. Bayesian networks were unable to detect direct and recurrent

connectivity patterns identified by DBN.

There are some limitations of the present method. First, the strengths of connections in networks derived from DBN are within the range $[0, 1]$, which could not differentiate between the negative and positive influence. Moreover, different orders (delay) of connection among regions may require high-order DBN rather than first-order Markovian models. However, such extension needs an efficient structure algorithms to account for variability of the delays and estimating the increased number of parameters. Although possible intra-scan connectivity may exist in fMRI data due to low temporal sampling, current DBN model did not taken into account these instantaneous interactions as these bi-directional intra-scan connections do not reflect the causal relationships and affect the acyclic constraints of Bayesian networks. The present approach still assumes time-invariant effective connectivity among brain regions while more and more empirical evidence suggests that the connectivity between brain areas may be time-variant. This calls for the extension of current DBN framework to model the changes of network structures along time. The proposed DBN approach is able to include all regions in the brain atlas. The main difficulty of the structure learning algorithm in DBN framework is that it is NP-hard problem to find the optimal structure. For the convergence of MCMC sampling algorithm, substantial amounts of computing time and memory space are required to search the huge sample space, which is computationally prohibitive. Further investigation of efficient structure learning algorithm for large networks in DBN could be a worthwhile pursuit. Nevertheless, DBN showed promise in estimating connectivity structure taking into consideration of temporal and causal relationship as evidenced by simulations on synthetic data and two experiments with real data. It is wise to combine confirmatory and exploratory techniques to decompose brain connectivity in a more efficient manner. The learned structure from training data could also be used as a significant feature to distinguish different groups of subjects from the test database (unknown labels).

Because of the high complexity of human brain networks, functional integration studies still have a long way to go. The full interpretation of the effective connectivity model derived from exploratory method remains difficult. Furthermore, although fMRI is able to measure mental activity both spatially and temporally, temporal information at the neuronal level may be limited to the hemodynamic level because of the temporal smoothing inherent in neuronal and hemodynamic coupling [43]. EEG and MEG studies, however, directly measure neural activity of the brain at a temporal resolution in the order of milliseconds. Therefore, it is essential to fuse fMRI and EEG together for accurate study of functional integration and causality. On the other hand, an estimate of anatomical connectivity in white matter of human brain can be obtained from measurements of the diffusion tensor image (DTI). A fusion of fMRI and DTI may serve as constraints for modelling brain connectivity. The multimodality approach should lead to more reliable inferences regarding anatomical and functional connectivity of the brain than those made based on a single modality.

Chapter 6

Conclusion

6.1 Summary and Discussion

A key objective in neuroscience is to improve our understanding of the relationship between structure and function of the human brain. Magnetic resonance imaging (MRI), which produces both structure and functional MR images, is a magnificent tool to investigate the human brain. Besides expert knowledge, computer technology is used to automatically discover useful knowledge from the two kinds of MR images to help clinical diagnosis and treatment planning. For example, quantitative study of brain structures from structural MRI requires fast and accurate segmentation of brain structures. Manual segmentation in clinical study cannot cope with the growing size and number of images. Over the past decade, researchers have developed various psychological tasks to localize functional specificity of the brain from BOLD signal changes in functional MRI. In addition, the study of brain connectivity has recently become the focus of neuroscience community, which provides exciting insights to the holistic mechanisms of human mental processes.

Because comprehensive brain studies using MRI/fMRI usually involve three closely related aspects: structures, activations, and connectivity. Accurate detection of functional activation from fMR images depends on localization of activation, i.e., segmen-

tation of structural MR images. Then structural and functional information act as the prior knowledge for brain connectivity inference. All these aspects involve individual variability and little previous work has been done to address this issue. Therefore, to this end, we have contributed and developed novel techniques for (1) identifying brain sub-cortical structures from MRI; (2) detecting functional activation from fMRI; and (3) learning effective brain connectivity from fMRI, that could explicitly take into account the individual variability of the brain.

Though multiple techniques have been applied to the segmentation of sub-cortical structures in structural MRI, because of their inability to address issues like intensity inhomogeneities, low contrast, and structural variability within a single framework, most segmentation methods only generate a coarse location of the desired objects. In this thesis, we developed a novel algorithm for the segmentation of sub-cortical structures that accounts for, among many other factors, the inter-subject structural variability. A set of fuzzy templates of structures based on their features are first created from a set of training images by defining the fuzzy membership functions and fusing the feature information. Segmentation is then performed by combining the registered fuzzy templates of structures and tissue maps of the test image. The final defuzzyfication optimizes the certainty in intensity, location, relative position, and tissue content of the structure. Unlike previous approaches, the method does not require specific expert definitions of each structure or manual interactions during the segmentation process, instead, the prior knowledge of structures are automatically extracted from pre-segmented images during the training phase. Moreover, the fuzzy templates are capable of modelling structural variability among subjects. The technique was demonstrated on the segmentation of five structures: thalamus, putamen, caudate, hippocampus, and amygdala, from a group of MR images. The performance of the method was comparable or superior to previous techniques. It can be easily extended to the detection of other structures or cortical areas.

Previous hypothesis-driven approaches for the detection of activation from fMRI data assume *a prior* knowledge of activation pattern and a specific structure for noise, and thus are not suitable for detecting complicated or unknown activation responses with artifacts. Data-driven methods such as PCA, ICA, and clustering are capable of discovering activation from fMRI data without prior assumption, but they suffer from the difficulty in interpreting the results, i.e., the discovered clusters do not have clear physiological meanings. In this thesis, we proposed two fuzzy approaches: fuzzy feature modelling (FFM) and weighted contextual fuzzy c-means clustering (wcFCM), both approaches are based on a novel feature space extracted from fMRI time-series using a sequence of temporal-sliding-windows (TSWs) in an automated manner. In FFM, a fuzzy model based on the extracted features is first developed by a gradient descent method and the fuzzy maps are incrementally updated. In wcFCM, the minimization of the weighted contextual cost function allows the incorporation of contextual constraints adaptively. Both methods do not assume any prior structures for signal or noise and do not have difficulty in relating the results to neurophysiology. Present methods outperformed statistical parametric mapping (SPM) approach in detecting activation from a group of synthetic fMRI data contaminated by both independent and correlated noise. Furthermore, the effect of the variability of HRF was investigated. The experimental results showed that both FFM and wcFCM were capable of adapting to the changes of HRF among individuals. In addition, wcFCM was capable of incorporating neighborhood information adaptively in the detection process, which further increased the detection accuracy. The methods were also validated on real fMRI datasets; the detected activations were consistent with evidence from neuroscience and were similar to those obtained from SPM.

Dynamic Bayesian networks (DBN) was proposed to learn the structure of effective brain connectivity from fMRI data in an exploratory manner. The proposed approach does not require any prior model and a MCMC-simulation based structure learning

method is used to find an optimal solution in the global sense. Unlike Bayesian networks (BN), DBN is capable of capturing temporal relationships among brain regions to infer causal effects as well as modelling inter-subject variability in effective connectivity. Synthetic fMRI data were generated to demonstrate the method, in which the effects of hemodynamic noise, scanner noise, and sampling in fMRI experiments were evaluated. The method had comparable performance as Granger causal mapping (GCM). The connectivity models inferred from two sets of real fMRI data by the present method were consistent with the previous literature and more informative than those discovered by BN. The present method is limited to first-order temporal precedence, which could be extended to model higher-order temporal relationships and time-variant connectivity structures. An extension of DBN was also proposed to handle the inter-subject variability of effective connectivity.

6.2 Future Work

The results achieved in detecting brain structures, activation, and connectivity by handling their variability in this thesis are very promising. Because of the close relationships and dependency among brain structure and functions, the methods proposed in this thesis could be integrated and extended in many directions.

6.2.1 Segmentation of Brain Structures

The proposed method is an attractive way to segment subcortical structures from MR images using fuzzy logic and information fusion. Fuzzy templates of each subcortical structure are first constructed during training stage, and then registered back to a test image. After incorporating tissue fuzzy maps, a defuzzification step is used to decide on final shape of the structure. The accuracy of tissue segmentation and

image registration largely influences the performance of subcortical structure segmentation. A more accurate registration and tissue segmentation algorithm is required. Multi-objective optimization approach could be developed to achieve structure segmentation, tissue segmentation, and registration simultaneously.

Moreover, the proposed approach should be extended to segment cortical brain regions. However, because of greater variability in shape (cortex folding) and size of cortical brain regions, it is possible to combine the present approach with deformable models. In addition, more brain areas, both cortical and subcortical, could be included to the relative spatial relationships of the target structure. As a result, a precise and complete delineation of the relationships among all brain regions are available, which could potentially improve segmentation accuracy.

6.2.2 Detection of Activation from fMRI

Despite of the proposed approaches, there are still several issues in the detection of brain activations from fMRI data. Firstly, the normalization step in preprocessing of fMRI time-series creates uncertainty and artifacts in group study. The variability of HRF exists among brain regions as well as individual subjects. Moreover, functional activation results of group analysis is simply overlaid on MR structure images and only Talairach and MNI coordinates are reported. Scientists usually use some softwares or reference atlas to determine the exact location and structure of the activated areas and some voxels may be named wrongly during the process.

Because the functional activation of the brain is mostly located in gray matter, little on white matter, and none on CSF, the anatomical information, e.g., fuzzy templates of brain structures and tissue priors, could be incorporated into the process of fMRI analysis. The precise anatomical information enables differentiation of HRF properties as well as removal of false activated voxels. Intelligent data fusion approaches for

structural and functional MRI are capable of increasing the system sensitivity to complex or unknown activation responses and helping to determine precise locations of activation.

6.2.3 Inference of Effective Connectivity

When analyzing effective connectivity among brain regions, different orders (delay) of connections among regions may require high-order DBN rather than the first-order Markovian model. The most difficult part is to develop an efficient structure leaning algorithm to account for variability of the delay. The present approach still assumes time-invariant effective connectivity among brain regions while more and more empirical evidence suggests that the connectivity among brain areas may be time-variant. This calls for the extension of current DBN framework to model the changes of network structures along time.

The study of anatomical connectivity of the human brain is able to validate the effective connectivity found from fMRI data. An estimate of anatomical connectivity in white matter is obtained by fibre tracking algorithms applying on diffusion tensor images (DTI). A Monte-Carlo simulation method was proposed to determine the probability that a virtual particle diffuses between two points, with the probability of a jump from a given voxel in a particular direction, based on the local value of diffusion tensor components [41]. The fusion of fMR and DT images may serve as the future for modelling effective connectivity. The combined modality approach will render more reliable information regarding the anatomical and functional connectivity of the brain than methods based on a single modality.

6.2.4 Applications to Studies of Brain Disease

Quantitative analysis of brain structures in MRI is a fundamental issue in the assessment of structural abnormalities of the brain [184]. Brain diseases like schizophrenia and dementia often require high precision detection of subtle volumetric differences of brain regions between patients and healthy control subjects. Similarly, there exist differences of functional and anatomical connectivity among normal subjects and patients. The differences in connectivity or dis-connectivity patterns give fruitful and more insights to underlying mechanisms of brain disease.

This thesis explored novel techniques for analysis of both structural and functional MR images for advanced functional brain studies. The approaches were developed for: (1) segmentation of subcortical structures from structural MR images, (2) detection of activation from functional MRI images, and (3) inference of brain connectivity. More importantly, this thesis research into individual variability of brain structures, function, and connectivity in a comprehensive manner. This work could be extended so that subjects can be automatically identified belonging to different categories such as patients with specific brain disease or healthy subjects by quantifying the differences of brain anatomy, function, and connectivity.

Appendix A: Derivation of Fuzzy Maps

With reference to section 3.2, the detailed derivations of fuzzy feature maps are as follows:

We model the intensity characteristics of the structure l in the normalized training image $\tilde{f}^j$ as a fuzzy set, whose fuzzy membership function is estimated from the intensity histogram of G_j^l as defined in Eq. 3.1. Figure 1 shows a typical fuzzy membership function of intensity for left thalamus in one training image. The rational behind this definition is to differentiate largely between those voxels belonging to a particular structure and other voxels, and to give higher fuzzy memberships to more prominent voxels and lower values to less prominent ones, based on the intensity distribution. For spatial location, fuzzy membership function is estimated from histograms of three dimensional values in the similar manner as shown in Figure 1, resulting in Eq. 3.2.

For relative spatial relations, both relative distance and relative direction of the target structure with respect to other structures are considered. The equations in section 3.2.1 can be illustrated by Figure 2. A vector marked in red is formed from the centroid of one reference structure to each voxel i in the target structure. With reference to Eqs. 3.4 to 3.7, the distance between the two points and the angle of this vector with respect to xy-plane, xz-plane, and yz-plane are denoted in the graph.

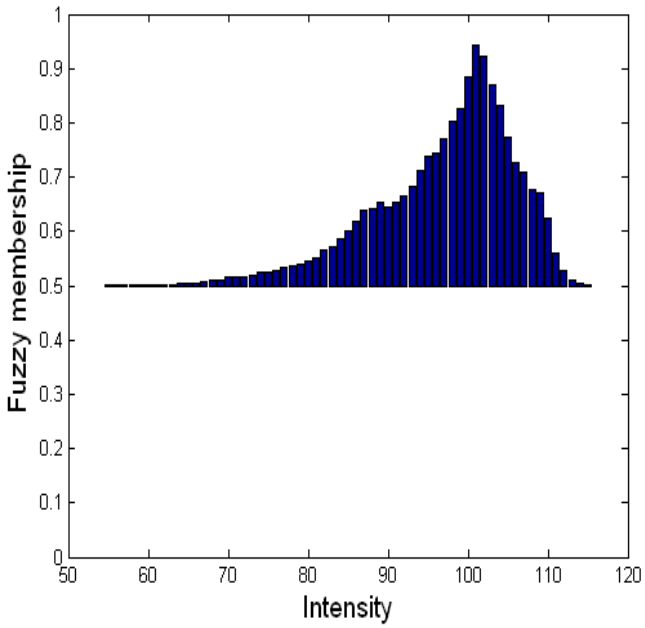


Figure 1 Fuzzy membership function of intensity left thalamus

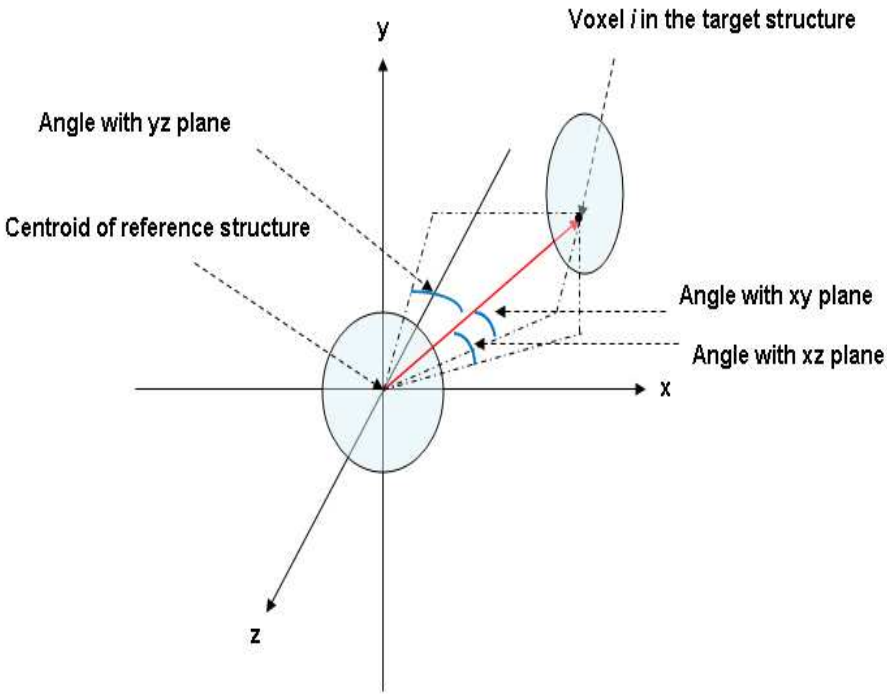


Figure 2 Illustration of derivation of fuzzy memberships of relative spatial relationships among structures.

References

[1] B. J. Casey, M. Davidson, and B. Rosen. Functional magnetic resonance imaging: basic principles of and application to developmental science. *Developmental Science*, 5(3):301–309, 2002.

[2] M. Brett, I. S. Johnsrude, and A. M. Owen. The problem of functional localization in the human brain. *Nature Reviews*, 3:243–249, 2002.

[3] J. Talairach and P. Tournoux. Co-planar stereotaxic atlas of the human brain. *Thieme Medical Publisher*, January 1988.

[4] M. Sonka and J. M. Fitzpatrick, editors. *Handbook of medical imaging*, volume 1 and 2 Medical image processing and analysis. SPIE Press, 2000.

[5] D. L. Pham, C. Xu, and J. L. Prince. Current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2:315–338, 2000.

[6] A. Meyer-Baese, A. Wismueller, and O. Lange. Comparison of two exploratory data analysis methods for fMRI: unsupervised clustering versus independent component analysis. *IEEE Transactions of Information Technology in Biomedicine*, pages 387–398, September, 2004.

[7] K. J. Worsley and K. J. Friston. Analysis of fMRI time-series revisited-again. *Neuroimage*, 2:173–181, 1995.

[8] P. McCullagh and J. A. Nelder. *Monographs on statistics and applied probability*, chapter Generalized Linear Models. Chapman and Hall, London, 2nd edition, 1989.

[9] S. Ngan and X. Hu. Analysis of functional magnetic resonance imaging data using self-organizing mapping with spatial connectivity. *Magnetic Resonance in Medicine*, 41:939–946, 1999.

[10] K. J. Friston, C. D. Frith, and R. S. J. Frackowiak. Time-dependent changes in effective connectivity measured with PET. *Human Brain Mapping*, 1:69–80, 1993.

[11] A. R. McIntosh and F. Gonzalez-Lima. Structural modeling of functional neural pathways mapped with 2-deoxyglucose: effects of acoustic startle habituation of the auditory system. *Brain Research*, 547:295–302, 1991.

[12] K. J. Friston. Dynamic causal modeling. *NeuroImage*, 19:1273–1302, 2003.

[13] E. Bullmore, B. Horwitz, G. Honey, M. Brammer, S. Williams, and T. Sharma. How good is good enough in path analysis of fMRI data? *NeuroImage*, 11:289–301, 2000.

[14] R. Goebel, A. Roebroecka, D. Kimb, and E. Formisanoa. Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modelling and granger causality mapping. *Magnetic Resonance Imaging*, 21:1251–1261, 2003.

[15] X. Zheng and J. C. Rajapakse. Learning functional structure from fMRI images. *Neuroimage*, 31(4):1601–1613, 2006.

[16] J. R. Hesselink. Basic principles of MR imaging. http://spinwarp.ucsd.edu/NeuroWeb/Text/br-100.htm, 2004.

[17] K. A. Johnson. Neuroimaging Primer. http://www.med.harvard.edu/AANLIB/, 2004.

[18] Joseph P. Hornak. The basics of MRI. http://www.cis.rit.edu/htbooks/mri/, 2007.

[19] S. Ogawa, T. M. Lee, and B. Barrere. The sensitivity of magnetic resonance image signals of a rat brain to changes in the cerebral venous blood oxygenation. 29:205–210, 1993.

[20] S. Ogawa, T. M. Lee, A. R. Kay, and D. W. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. 87:9868–9872, 1990.

[21] N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann. Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412:150–157, 2001.

[22] R. S. Menon and S. G. Kim. Spatial and temporal limits in cognitive neuroimaging with fMRI. *Trends in Cognitive Sciences*, 3:207–216, 1999.

[23] P. Mansfield. Multi-planar image formation using NMR spin echoes. *J. Phys. C.*, 10:580–594, 1977.

[24] S. A. Huettel, A. W. Song, and G. McCarthy. *Functional Magnetic Resonance Imaging*. Sinauer Associates, Inc., 2004.

[25] J. C. Rajapakse, F. Kruggel, J. M. Maisog, and D. Y. von Cramon. Modeling hemodynamic response for analysis of functional MRI time-series. *Human Brain Mapping*, 6:283–300, 1998.

[26] Department of Kinesiology and University of Texas at Austin Health Education. Brain. http://www.edb.utexas.edu/syllabus/farrar/lectures/brain.html, 1999.

[27] E. H. Chudler. Explore the nervous system. http://faculty.washington.edu/chudler/introb.html, 2004.

[28] G. Wideman. Orientation and voxel-order terminology: RAS, LAS, LPI, RPI, XYZ and all that. http://wideman-one.com/gw/brain/analyze/orientterms.htm, June 2003.

[29] Internation Consortium of Brain Mapping. ICBM single subject MRI brain template. http://www.loni.ucla.edu/ICBM/ICBM_BrainTemplate.html, 2004.

[30] MNI space: Cognition and Brainscience Unit Imaging home page. The MNI brain and the Talairach atlas. http://www.mrc-cbu.cam.ac.uk/Imaging/mnispace.html, 2004.

[31] J. Mazziotta, A. Toga, A. Evans, P. Fox, and J. Lancaster. A probalistic atlas of the human brain: theory and rationale for its development. *NeuroImage*, 2:89–101, 1995.

[32] J. Mazziotta, A. Toga, A. Evans, P. Fox, J. Lancaster, K. Zilles, R. Woods, T. Paus, G. Simpson, B. Pike, and et al. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Phil. Trans. of the Royal Society*, 356:1293–1322, 2001.

[33] A. Hammers, R. Allom, M. J. Koepp, S. L. Free, R. Myers, L. Lemieux, T. N. Mitchell, D. J. Brooks, and J. S. Duncan. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Human Brain Mapping*, 19:224–247, 2003.

[34] A. Hammers, M. J. Koepp, S. L. Free, M. Brett, M. P. Richardson, C. Labbe, V. J. Cunningham, D. J. Brooks, and J. Duncan. Implementation and application of a brain template for multiple volumes of interest. *Human Brain Mapping*, 15:165–174, 2002.

[35] S. Finger. *The origins of neuroscience: a history of explorations into brain function.* New York: Oxford University Press, 1994.

[36] R. S. J. Frackowial, K. J. Friston, C. D. Frith, R. J. Dolan, and J. C. Mazziotta. *Human Brain Function.* Academic Press, San Diego, 1997.

[37] M. M. Mesulan. Fron sensation to cognition. *Brain*, 121:1013–1052, 1998.

[38] O. Sporns. Network analysis, complexity, and brain function. *Complexity*, 8:56–60, 2002.

[39] The free encyclopedia Wikipedia. Neuron. http://en.wikipedia.org/wiki/Neuron, 2007.

[40] G. M. Shepherd, editor. *The synaptic organization of the brain.* New York: Ocford University Press, 1998.

[41] M. A. Koch, D. G. Norris, and M. Hund-Georgiadis. An investigation of functional and anatomical connectivity using magnetic resonance imaging. *NeuroImage*, 16:241–250, 2002.

[42] K. J. Friston, C. D. Frith, P. F. Liddle, and R. S. J. Frackowiak. Functional connectivity: the principle component analysis of large (PET) data sets. *Journal of Cerebral Blood Flow and Metabolism*, 13:5–14, 1993.

[43] L. Lee, L. M. Harrison, and A. Mechell. A report of the functional connectivity workshop, (Dusseldorf 2002). *NeuroImage*, 19:457–465, 2003.

[44] T. M. Cover and J. A. Thomas. *Elements of information theory.* Wiley, New York, 1991.

[45] L. Lemieux, A. Hammers, T. Mackinnon, and R. S. N. Liu. Automatic segmentation of the brain and intracranial cerebrospinal fluid in T1-weighted volume

MRI scans of the head and its application to serial cerebral and intracranial volumetry. *Magnetic Resonance in Medicine*, 49(5):872–884, 2003.

[46] V. Barra and J. Boire. Automatic segmentation of subcortical brain structures in MR images using information fusion. *IEEE Transactions on Medical Imaging*, 20(7):549–558, July 2001.

[47] D. L. Collins, A. P. Zijdenbos, W. F. C. Baare, and A. C. Evans. Animal+insect: improved cortical structure segmentation. *Proceeding of Information Processing in Medical Imaging (IPMI)*, pages 210–223, 1999.

[48] J. G. Sled and G. B. Pike. Understanding intensity nonuniformity in MRI. *Proc. of Medical Image Computing Computer-Assisted Intervention (MICCAI), Berlin, Germany, Lecture notes in Computer Science*, 1496:614–622, 1998.

[49] Y. J. Zhang. A survey of evaluation methods for image segmentation. *Pattern Recognition Letters*, 29:1335–1346, 1996.

[50] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int. J. Comput. Vis.*, 1:321–331, 1988.

[51] J. A. Sethian. *Level set methods and fast marching methods.* Cambridge University Press, U. K., 1999.

[52] C. Ciofolo, C. Barillot, and P. Hellier. Combining fuzzy logic and level set methods for 3D MRI brain segmentation. *Proceedings of IEEE International Conference of Biomedical Imaging*, pages 161–164, May 2004.

[53] Y. Fan and T. Jiang anf D. J. Evans. Volumetric segmentation of brain images using parallel genetic algorithms. *IEEE Transactions on Medical Imaging*, 21(8):904–910, 2002.

[54] C. J. Yang, H. D. Tagare, L. H. Staib, and J. S. Duncan. Segmentation of 3D deformable objects with level set based prior models. *Proc. of IEEE International Conference of Biomedical Imaging*, pages 85–88, May 2004.

[55] D. V. Iosifescu, M. E. Shenton, S. K. Warfield, R. Kikinis, J. Dengler, F. A. Jolesz, and R. W. McCharley. An automated registration algorithm for measuring MRI subcortical brain structures. *Neuroimage*, 6:13–25, 1997.

[56] A. Kelemen, G. Szekely, and G. Gerig. Elastic model-based segmentation of 3D neuoradiological data sets. *IEEE Transactions on Medical Imaging*, 18:828–839, 1999.

[57] X. Han and B. Fischl. Atlas renormalization for improved brain MR image segmentation across scanner platforms. *IEEE Transactions on Medical Imaging*, 26(4):479–486, 2007.

[58] B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A. M. Dale. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33:341–355, 2002.

[59] J. Xue, S. Ruan, and B. Moretti. Fuzzy modeling of knowledge for MRI brain structure segmentation. *ICIP*, I:617–620, 2000.

[60] V. Barra and J. Boire. A general framework for the fusion of anatomical and functional medical images. *NeuroImage*, 13:410–424, 2001.

[61] L. Amini, H. Soltanian-Zadeh, C. Lucas, and M. Gity. Automatic segmentation of thalamus from brain MRI integrating fuzzy clustering and dynamic coutours. *IEEE Transactions on Biomedical Engineering*, 51(5):800–811, 2004.

[62] O. Colloit, O. Camara, and I. Bloch. Integration of fuzzy spatial relations in

deformable models – application to brain MRI segmetentation. *Pattern Recognition*, 39:1401–1414, 2006.

[63] A. J. Worth, N. Makris, M. R. Patti, J. M. Goodman, E. A. Hoge, V. S. Caviness, and D. N. Kennedy. Precise segmentation of the lateral ventricles and caudate nucleus in MR brain images using anatomically driven histograms. *IEEE Transactions on Medical Imaging*, 17(2):303–309, 1998.

[64] V. A. Magnotta, D. Heckel, A. C. Andreasen, T. Cizaldo, P. W. Corson, J. C. Echrhardt, and W. T. Yuh. Meansurement of brain structures with artificial neural networks: two- and three-dimensional applications. *Radiology*, 211:781–790, 1999.

[65] S. Powell, V. Magnotta, H. Johnson, and N. Andreasen. Automated brain segmentation using neural networks. *Medical Imaging, Proceedings of SPIE*, 6144:1–10, 2006.

[66] M. Sonka, S. K. Tadikonda, and S. M. Collins. Knowledge-based interpretation of MR brain images. *IEEE Transactions on Medical Imaging*, 15(4):443–452, 1996.

[67] M. Chupin, A. R. Mukuna-Bantumbakulu, D. Hasboun, E. Bardinet, S. Baillet, S. Kinkingnehun, L. Lemieux, B. Dubois, and L. Garnero. Anatomically constrained region deformation for automated segmentatio of the hippocampus and the amygdala: Method and validation on controls and patients with Alzheimer's disease. *NeuroImage*, 34:996–1019, 2007.

[68] Y. Xia, K. Bettinger, L. Shen, and A. L. Reiss. Autoamtic segmentation of the caudate nucleus from human brain MR images. *IEEE Transactions on Medical Imaging*, 26(4):509–517, 2007.

[69] G. A. Hossein-Zadeh, H. Soltanian-Zadeh, and B. A. Ardekani. Multiresolution fMRI activation detection using translation invariant wavelet transform and statistical analysis based on resampling. *IEEE Transactions on Medical Imaging*, 22(3):302–314, March 2003.

[70] K. J. Friston, J. Ashburner, C. D. Frith, J-B Poline, J. D. Heather, and R. S. J. Frackowiak. Spatial registration and normalization of images. *Human Brain Mapping*, 2:165–189, 1995.

[71] J. Ashburner. *Computational Neuroanatomy*. PhD thesis, University College London, 2000.

[72] R. P. Woods, S. T. Grafton, C. J. Holmes, S. R. Cherry, and J.C. Mazziotta. Automated image registration: I. general methods and intrasubject, intramodality validation. *J. Comput. Assist. Tomogr.*, 22(1):139–152, 1998.

[73] J. Ashburner and K. Friston. Voxel-based morphometry - the methods. *NeuroImage*, 11:805–821, 2000.

[74] K. J. Friston, S. William, R. Howard, R. S. J. Frackowiak, and R. Turner. Movement related effects in fMRI time-series. *Magnetic Resonance in Medicine*, 35:346–355, 1996.

[75] J. L. R. Andersson, C. Hutton, J. Ashburner, R. Turner, and K. Friston. Modelling geometric deformations in EPI time series. *NeuroImage*, 13:903–919, 2001.

[76] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multi-modality image registration based on information theory. *Information Processing in Medical Imaging. Dordrecht, The Netherlands: Kluwer Academic Publishers*, pages 263–274, 1995.

[77] C. M. WellsIII, P. Viola, H. Atsumi, and S. Nakajima. Multi-modal volume

registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35–51, 1996.

[78] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, 1997.

[79] C. Studholme, D. L. G. Hill, and D. J. Hawkes. A normalized entropy measure of 3D medical image alignment. *Proc. of Medical Imaging, San Diego, USA*, 3338:132–143, 1998.

[80] F. Maes, D. Vandermeulen, and P. Suetens. Medical image registration using mutual information. *Proceedings of the IEEE*, 91(10):1699–1722, 2003.

[81] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging*, 22(8):986–1004, 2003.

[82] J. Ashburner, P. Neelin, D. L. Collins, A. Evans, and K. Friston. Incorporating prior knowledge into image registration. *NeuroImage*, 6:344–352, 1997.

[83] P. T. Fox. Spatial normalization origins: objectives, applications, and alternatives. *Human Brain Mapping*, 3:161–164, 1995.

[84] J. Ashburner and K. Friston. Multimodal image coregistration and partitioning - a unified framework. *NeuroImage*, 6(3):209–217, 1997.

[85] J. Ashburner and K. Friston. Nonlinear spatial normalization using basis functions. *Human Brain Mapping*, 7(4):254–266, 1999.

[86] P. Bandettini, A. Jesmanowicz, E. Wong, and J. Hyde. Processing strategies for time-course data sets in functional MRI of human brain. *Magnetic Resonance Imaging*, 30:161–173, 1993.

[87] K. Friston. *Human brain function*, chapter Introduction: experimental design and statistical parametric mapping. Academic Press, 2nd edition, 2003.

[88] K. J. Friston, A. P. Holmes, K. J. Worsley, J.-B. Poline, C. D. Frith, and R. S. J. Frackowiak. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2:189–210, 1995.

[89] Wellcome Department of Imaging Neuroscience. Statistical Parametric Mapping. http://www.fil.ion.ucl.ac.uk/spm/, 2004.

[90] K. J. Friston, A. P. Holmes, J.-B. Poline, P. J. Grasby, SCR Williams, R. S. J. Frackowiak, and R. Turner. Analysis of fMRI time-series revisited. *Neuroimage*, 2:45–53, 1995.

[91] M. Brett, W. Penny, , and S. Kiebel. *Human brain function*, chapter Introduction to Random Field Theory. Academic Press, 2nd edition, 2003.

[92] R. J. Alder. *The Geometry of Random Fields*. New York: Wiley, 1981.

[93] K. J. Worsley. Local maxima and the expected euler characteristic of excursion sets of χ^2, F and t fields. *Advances in Applied Probability*, 26:13–42, 1994.

[94] K. J. Worsley, M. Andermann, T. Koulis, D. MacDonald, and A. C. Evans. Detecting changes in nonisotropic images. *Human Brain Mapping*, 8:98–101, 1999.

[95] K. J. Worsley, S.Marrett, P. Neelin, A. C. Vandal, K. J. Friston, and A. C. Evans. A united statistical approach for determing significant signals in images of cerebral activation. *Human Brain Mapping*, 4:58–73, 1996.

[96] K. J. Worsley, J.E. Taylor, F. Tomaiuolo, and J. Lerch. Unified univariate and multivariate random field theory. *Neuroimage*, 23:189–195, 2004.

[97] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57:289–300, 1995.

[98] C. R. Genovese, N. Lazar, and T. E. Nichols. Thresholding of statistical maps in functional neuroimaging using the False Discovery Rate. *NeuroImage*, 15:870–878, 2002.

[99] W. Backfrieder, R. Baumgartner, M. Samal, E. Moser, and H. Bergmann. Quantification of intensity variations in functional MR images using rotated principal components. *Physics Medicine Biology*, 41:1425–1438, 1996.

[100] M. J. McKeown, S. Makeig, G. G. Brown, T. P. Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski. Analysis of fMRI data by blind separation into independent spatial components. *Human Brain Mapping*, 6:160–188, 1998.

[101] B. Biswal and J. Ulmer. Blind source separation of multiple signal sources of fMRI data sets using independent component analysis. *Journal of Computer Assisted Tomography*, 23(2):265–271, 1999.

[102] V. Calhoun, T. Adali, G. Pearlson, and J. Pekar. Spatial and temporal independent component analysis of functional MRI data containing a pair of task related waveforms. *Human Brain Mapping*, 13:43–253, 2001.

[103] W. Lu and J. C. Rajapakse. Approach and applications of constrained ICA. *IEEE Transactions on Neural Networks*, 16(1):203–212, 2005.

[104] J. V. Stone, J. Porrill, N. R. Porter, and I. D. Wilkinson. Spatiotemporal independent component analysis of event-related fMRI data using skewed probability density functions. *NeuroImage*, 15(2):407–421, 2002.

[105] K. Chuang, M. Chiu, C. Lin, and J. Chen. Model-free functional MRI analysis

using Kohonen clustering neural network and fuzzy c-means. *IEEE Transactions on Medical Imaging*, 18:1117–1128, 1999.

[106] J. C. Rajapakse and J. Piyaratna. Bayesian approach to segmentation of statistical parametric maps. *IEEE Transactions on Biomedical Engineering*, 48:1186–1194, 2001.

[107] Y. Wang and J. C. Rajapakse. Contextual modeling of functional MR images with conditional random fields. *IEEE Transactions on Medical Imaging*, 25(6):804–812, 2006.

[108] V. V. Tscharner and K. R. Thulborn. Specified-resolution wavelet analysis of activation patterns from BOLD contrast fMRI. *IEEE Transactions on Medical Imaging*, 20(8):704–714, 2001.

[109] M. W. Woolrich, M. jenkinson, J. M. Brady, and S. M. Smith. Fully bayesian spatio-temporal modeling of fMRI data. *IEEE Transactions on Medical Imaging*, 23(2):213–231, 2004.

[110] A. L. Vazquez and D. C. Noll. Nonlinear aspects of the BOLD response in functional MRI. *Human Brain Mapping*, 40:249–260, 1998.

[111] E. Formisano and R. Goebel. Tracking cognitive processes using fMRI mental chronometry. *Curr. Opin. Neurobiol.*, 13:174–181, 2003.

[112] D. Bavelier, A. Tomann, C. Hutton, T. Mitchell, D. Corina, G. Liu, and H. Neville. Visual attention to the periphery is enhanced in congenitally deaf individuals. *The Journal of Neuroscience*, 20:RC93 1–6, 2000.

[113] A. R. McIntosh and F. Gonzalez-Lima. Structural equation modeling and its application to network analysis in functional brain imaging. *Human Brain Mapping*, 2:2–22, 1994.

[114] A. R. McIntosh, C. L. Grady, L. G. Ungerieider, J. V. Haxby, S. I. Rapoport, and B. Horwitz. Network analysis of cortial visual pathways mapped with PET. *The Journal of Neuroscience*, 14(2):655–666, 1994.

[115] B. J. Krause, B. Horwitz, J. G. Taylor, D. Schmidt, F. M. Mottaghy, H. Herzog, U. Halsband, and H. W. Muller-Gartner. Network analysis in episodia encoding and retrieval of word-pair associates: A PET study. *The Journal of Neuroscience*, 11:3293–3301, 1999.

[116] K. A. McKiernan, L. L. Conant, A. Chen, and J. R. Binder. Development and cross-validation of a model of linguistic processing using neural network and path analyses with fMRI data. *NeuroImage*, 13(6):2–2, 2001.

[117] R. Nezafat, R. Shadmehr, and H. H. Holcomb. Long-term adaptation to dynamics of reading movement: A PET Study. *Exp. Brain Res.*, 140:66–76, 2001.

[118] K.L. Petersson, A. Reis, S. Askelof, A. Castro-Caldas, and M. Ingvar. Language processing modulated by literacy: a network analysis of verbal repetition in literate and illiterate subjects. *Congnitive Neuroscience*, pages 364–382, 2000.

[119] G. D. Honey, C. H. Y. Fu, J. Kim, M. J. Brammer, T. J. Groudace, J. Suckling, E. M. Pich, and S. C. R. William. Effects of verbal working memory load on corticocortial connectivity modeled by path analysis of functional magnetic resonace imaging data. *NeuroImage*, 17:573–582, 2002.

[120] L. Nyberg, A. R. McIntosh, R. Cabeza, L. Nilsson, S. Houle, R. Habib, and E. Tulving. Network analysis of Positron Emission Tomography regional cerebral blood flow data: Ensemble inhibition during episodic memory retrieval. *The Journal of Neuroscience*, 16(11):3753–3759, 1996.

[121] A. Mechelli, W. D. Penny, C. J. Price, D. R. Gitelman, and K. J. Friston.

Effective connectivity and intersubject variability: using a multisubject network to test differences and commonalities. *NeuroImage*, 17:1459–1469, 2002.

[122] K. A. Bollen. *Structural Equations with Latent Variables*. John Wiley and Sons, Inc., 1989.

[123] W. D. Penny, K. E. Stephan, A. Mechelli, and K. J. Friston. Modelling functional integration: a comparison of structural equation and dynamic causl models. *NeuroImage*, 23:264–274, 2004.

[124] W.D. Penny, Z. Ghahramani, and K.J. Friston. Bilinear dynamical systems. *Phil. Trans. R. Soc. B*, 360:983–993, 2005.

[125] J. Geweke. Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association*, 77:304–313, 1982.

[126] M. Eichler. A graphical approach for evaluating effective connectivity in neural systems. *Phil. Trans. R. Soc. B*, 360:953–967, 2005.

[127] P. A. Valdes-Sosa, J. M. Sanchez-Bornot, A. Lage-Castellanos, M. Vega-Hernandez, J. Bosch-Bayard, L. Melie-Garcia, and E. Canales-Rodriguez. Estimating brain functional connectivity with sparse multivariate autoregression. *Phil. Trans. R. Soc. B.*, 360:969–981, 2005.

[128] A. Roebroeck, E. Formisano, and R. Goebel. Mapping directed influence over the brain using granger causality and fMRI. *NeuroImage*, 25:240–242, 2005.

[129] J. C. Rajapakse, J. N. Giedd, and J. L. Rapoport. Statistical approach to segmentation of single-channel cerebral MR images. *IEEE Transactions on Medical Imaging*, 16:176–186, 1997.

[130] A. Pitiot, H. Delingette, P. M. Thompson, and N. Ayache. Expert knowledge-guided segmentation system for brain MRI. *Neuroimage*, 23:85–96, 2004.

[131] J. F. Mangin, F. Poupon, E. Duchesnay, D. Riviere, A. Cachia, D. L. Collins, A. C. Evans, and J. Regis. Brain morphometry using 3D moment invariants. *Medical Image Analysis*, 8:187–196, 2004.

[132] V. Barra, J. Lemarie, F. Durif, and J. Boire. Segmentation of the subthalamic nucleus in MR images using information fusion - a preliminary study for a computed-aided surgery of Parkinson's disease. *Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 1183–1184, 2001.

[133] The Center for Morphometric Analysis at Massachusetts General Hospital. MR brain images and their manual segmentations. http://www.cma.mgh.harvard.edu/ibsr/, 2004.

[134] R. Kikinis, M. E. Shenton, D. V. Iosifescu, R. W. McCarley, P. Saiviroonporn, H. H. Hokama, A. Robatino, D. Metcalf, C. G. Wible, C. M. Portas, R. M. Donnino, and F. A. Jolesc. A digital brain atlas for surgical planning, model driven segmentation, and teaching. *IEEE Visualizat. Comput. Graphics*, 2(3):232–241, 1996.

[135] J. C. Rajapakse and F. Kruggel. Segmentation with MR images with intensity inhomogenieties. *Image and vison computing*, 16:165–180, 1998.

[136] J. C. Rajapakse, C. DeCarli, A. Mclaughlin, J. N. Giedd, A. L. Krain, S. D. Hamburger, and J. L. Rapoport. Cerebral magnetic resonance image segmentation using data fusion. *J. Comput. Assist. Tomogr.*, 20(2):206–218, 1996.

[137] S. Ogawa, D. Tank, and R. Menon. Intrinsic signal changes accompanying sensory simularion: Functional brain mapping with magnetic resonance imaging. 89:5951–5955, 1992.

[138] E. Bullmore, C. Long, J. Fadili, G. Calvert, F. Zalaya, T. A. Carpenter, and M. Brammer. Colored noise and computational inference in neuophysiological (fMRI) time series analysis: Resampling mehtods in time and wavelet domains. *Human Brain Mapping*, 12:61–78, 2001.

[139] A. H. Andersen, D. M. Gash, and M. J. Avison. Principal component analysis of the dynamic response measured by fMRI. *Magnetic Resonance Imaging*, 17(6):795–815, 1999.

[140] T. J. Ross. *Fuzzy logic with engineering applications*. John Wiley and Sons Inc. 2nd edition, 2004.

[141] E. R. Cosman, J. W. Fisher, and W. M. Wells. Exact map activity detection in fMRI using a GLM with an ising spatial prior. *Proceedings of Medical Image Computing and Computer-Assisted Intervention*, 2:703–710, 2004.

[142] X. Descombes, F. Kruggel, and D. Y. von Cramon. Spatio-temporal fMRI analysis using Markov random fields. *IEEE Transactions on Medical Imaging*, 17(6):1028–1039, 1998.

[143] E. Salli, H. J. Aronen, S. Savolainen, A. Korvenoja, and A. Visa. Contextual clustering for analysis of functional MRI data. *IEEE Transactions on Medical Imaging*, 20(5):403–414, 2001.

[144] M. W. Woolrich, T. E. J. Behrens, C. F. Beckmann, and S. M. Smith. Mixture models with adaptive spatial regularization for segmentation with an application to fMRI data. *IEEE Transactions on Medical Imaging*, 24(1):1–11, 2005.

[145] D. L. Pham and J. L. Prince. An adaptive fuzzy segmentation algorithm for three-dimensional magnetic resonance images. *A. Kuba, M. Samal, and A. Todd-Pokropek (Eds.): Information processing in medical imaging (IPMI)*, pages 140–153, June–July 1999.

[146] M. N. Ahmed, S. M. Yamany, N. Mohamed, A. A. Farag, and T. Moriarty. A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data. *IEEE Transactions on Medical Imaging*, 21(3):193–199, 2002.

[147] Y. Yang, F. Zhang, C. Zheng, and P. Lin. Unsupervised image segmentation using penalized fuzzy clustering algorithm. *M. Gallagher, J. Hogan, and F. Maire (Eds): IDEAL 2005, LNCS*, 3578:71–77, 2005.

[148] S. Chen and D. Zhang. Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure. *IEEE Transactions on Systems, Man and Cybernetics*, 34(4):1907–1916, 2004.

[149] D. W. Jordan and P. Smith. *Mathematical techniques, an introduction for engineering, physical, and mathematical Sciences*. Addison-Wesley, 1996.

[150] G. James. *Modern engineering mathematics*. Oxford, U. K., Oxford Univ. Press, 1997.

[151] J. Bezdek. A convergence theorem for the fuzzy ISODATA clustering algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2:1–8, 1980.

[152] J. Bezdek, R. J. Hathaway, M. J. Sabin, and W. T. Tucker. Convergence theory for fuzzy c-means: counterexamples and repairs. *IEEE Transactions on Systems, Man, and Cybernetics*, 17:873–877, 1987.

[153] K. J. Friston, A. P. Holmes, and K. J. Worsley. How many subjects constitute a study? *Neuroimage*, 10:1–5, 1999.

[154] G. Bush, P. J. Whalen, B. R. Rosen, M. A. Jenike S. C. McInerney, and S. L. Rauch. The counting Stroop: An interference task specialized for functional neuroimaging validation study with functional MRI. *Human Brain Mapping*, 6:270–282, 1998.

[155] L. Tamm, V. Menon, C.K. Johnston, D.R. Hessl, and A.L. Reiss. fMRI study of cognitive interference processing in females with fragile X syndrome. *Journal of Cognitive Neuroscience*, 14(2):160–171, 2002.

[156] A. Mechelli, K. J. Friston, and C. J. Price. The effects of presentation rate during word and pseudoword read: A comparison of PET and fMRI. *Cognitive Neuroscience*, 12:145–156, 2000.

[157] J. C. Rajapakse, J. N. Giedd, C. DeCarli, J. W. Snell, A. McLaughlin, Y. C. Vauss, A. L. Krain, S. Hamburger, and J. L. Rapoport. A technique for single-channel MR brain tissue segmentation: Application to a pediatric sample. *J. Comput. Assist. Tomogr.*, 14(9):1053–1065, 1996.

[158] C. Buchel and K. J. Friston. Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated with structural equation modelling and fMRI. *Cerebral Cortex*, 7:768–778, 1997.

[159] J. C. Rajapakse, C. L. Tan, X. Zheng, S. Mukhopadyay, and K. Yang. Exploratory analysis of brain connectivity with ICA. *IEEE Engineering in Medicine and Biology Magazine*, 25(2):102–111, 2006.

[160] L. Harrison, W. D. Penny, and K. J. Friston. Multivariate autoregressive modelling of fMRI time series. *NeuroImage*, 19:1477–1491, 2003.

[161] D. M. Chickering. A transformational characterization of equivalent Bayesian network structures. *International Conference of Uncertainty in Artificial Intelligence (UAI)*, 11:87–98, 1995.

[162] L. Zhang, D. Samaras, N. Alia-Klein, N. Volkow, and R. Goldstein. Modeling neuronal interactivity using dynamic Bayesian networks. *Neural Information Processing Systems*, 2005.

[163] D. Husmeier. Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian network. *BioInformatics*, 19(17):2271–2282, 2003.

[164] N. Friedman, K. Murphy, and S. Russell. Learning the structure of dynamic probabilistic networks. *In G. F. Cooper and S. Moral (Eds.), Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI). Morgan Kaufmann Publishers, San Francisco, CA*, pages 139–147, 1998.

[165] N. Friedman. The Bayesian structural EM algorithm. *In G. F. Cooper and S. Moral (Eds.), Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI). Morgan Kaufmann Publishers, San Francisco, CA*, pages 129–138, 1998.

[166] S. Chib and E. Greenberg. Understanding the Metropolis-Hastings algorithm. *American Statistics*, 49:327–335, 1995.

[167] W. K. Hastings. Monte Carlo sampling methods using Markoc chains and their applications. *Biometrika*, 57:97–109, 1970.

[168] M. S. Goncalves, D. A. Hall, I. S. Johnsrudes, and M. P. Haggard. Can meaningful effective connectivities be obtained between auditory cortical regions? *NeuroImage*, 14:1353–1360.

[169] B. Horwitz, J. M. Rumsey, and B. C. Donohue. Functional connectivity of the angular gyrus in normal reading and dyslexia. *Proc. Natl. Acad. Sci. (USA)*, 95:8939–8944, 1998.

[170] E. R. Cosman and W. M. Wells. Bayesian population modeling of effective connectivity. *G. E. Christensen and M. Sonka (Eds.): Information Processing in Medical Imaging (IPMI)*, 3565:39–51, 2005.

[171] Dartmouth College fMRI DC. The fMRI data center. http://www.fmridc.org., 2004.

[172] K. P. Murphy. The Bayes net toolbox for MATLAB. *Computing Science and Statistics*, 33:331–351, 2001.

[173] D. Husmeier. Inferring dynamic Bayesian networks with MCMC. http://www.bioss.sari.ac.uk/∼dirk/software/DBmcmc/#Preliminaries., 2003.

[174] E. Bagarinao and S. Sato. Algorithm for vector autoregressive model parameter estimation using an orthogonalization procedure. *Annual Biomedical Engineering*, 30:260–271, 2002.

[175] R. Matsumoto, D. R. Nair, E. LaPresto, I. Najm, W. Bingaman, H. Shibasaki, and H. O. Luders. Functional connectivity in the human language system: a cortico-cortical evoked potential study. *Brain*, 127:2316–2330, 2004.

[176] B. Kolb and I. Q. Whishaw. *Fundamantal of Human Neuropsychology*. W. H. Freman and Company, 1986.

[177] M. I. Posner and G. J. DiGirolamo. *The Attentive Brain*, chapter Executive attention: conflict, target detection, and cognitive control, pages 401–423. Cambridge, MA: MIT Press, 1998.

[178] G. Hayward, G. M. Goodwin, C. J. Harmer, and R. W. McCharley. The role of the anterior cingulate cortex in the counting Stroop task. *Exp. Brain Res.*, 154:355–358, 2004.

[179] L. M. Shin, P. J. Whalen, R. K. Pitman, G. Bush, M. L. Macklin, N. B. Lasko, S. P. Orr, S. C. McInerney, and S. L. Rauch. An fMRI study of anterior cingulate function in posttraumatic stress disorder. *Society of Biological Psychaitry*, 0006–3223(01):01215–X, 2001.

[180] B. Faw. Pre-frontal executive committee for perception, working memory, attention, long-term memory, motor control, and thinking: A tutorial review. *Consciousness and cognition*, 12:83–139, 2002.

[181] C. J. Price. The anatomy of language: contributions from functional neuroimaging. *J. Anat.*, 179:335–359, 2000.

[182] K. G. Volz, R. I. Schubotz, and D. Y. von Cramon. Variants of uncertainty in decision-making and their neural correlates. *Brain Res. Bull.*, 67(5):403–412, 2005.

[183] D. Vanlancker, A. R. McIntosh, and S. Grafton. PET activation studies comparing two speech tasks widely used in surgical mapping. *Brain and Language*, 85:245–261, 2003.

[184] M. E. Shenton, R. Kikinis, F. A. Jolesz, S. D. Pollak, M. LeMay, C. G. Wible, H. Hokama, J. Martin, D. Metcalf, M. Coleman, and R. W. McCarley. Abnormalities of the left temporal lobe and thought disorder in schizophrenia: a quantitative magnetic resonance imaging study. *N. Engl. J. Med.*, 327:604–612, 1992.