

國立臺灣大學電機資訊學院電機工程學研究所



博士論文

Department of Electrical Engineering  
College of Electrical Engineering and Computer Science  
National Taiwan University  
Doctoral Dissertation

結構與功能性大腦磁振影像資訊擷取  
Information Extraction from Structural and Functional  
Brain MR Images using Binary Patterns

張哲維

Che-Wei Chang

指導教授：陳志宏 博士

Jyh-Horng Chen, Ph.D.

中華民國 103 年 1 月

January, 2014

國立臺灣大學博士學位論文  
口試委員會審定書

結構與功能性大腦磁振影像資訊擷取  
Information Extraction from Structural and Functional  
Brain MR Images using Binary Patterns

本論文係張哲維君（學號 F92921121）在國立臺灣大學電機工程學系完成之博士學位論文，於民國 103 年 1 月 21 日承下列考試委員審查通過及口試及格，特此證明。

口試委員：

陳志宏 (簽名)  
(指導教授)

陳建中

廖啟

林農禮

林奇德

張程

徐永年

江明毅

林仲治

曾明宇

郭立威

吳昌衛

系主任

廖婉芳 (簽名)

## 誌謝



能完成這篇論文，我要特別感謝我的指導教授陳志宏老師，因為老師多年來的耐心指導，在學術界宏大的見識與學養，以及對科學研究的熱誠，促使我能將一些有趣的想法化為對科學界的貢獻。此外，我也要感謝在學習機器學習理論與實作時，帶領台大團隊參與國際比賽的三位老師，林智仁教授、林守德教授以及林軒田教授。在該次比賽的團隊中，我獲益良多，並從過程中學習到了將機器學習的理論化為實作的方法與技巧。另外，我亦要感謝在學術上給我許多建議和幫助的前輩和老師們。首先要特別感謝台大電機系的何建璋博士，在研究和論文撰寫及投稿上提供非常多的幫助和指導。另外，感謝陽明大學的林慶波教授、國家衛生研究院的郭立威研究員、中央大學的吳昌衛教授、中山醫學大學的翁駿程教授，在學術研究的道路上提供的建言和幫助。同時，也真誠感謝所有口試委員的參與並提供我許多寶貴的意見。

另外，在此也要特別感謝實驗室裡一起研究及奮鬥的同學及學弟妹們。感謝已經畢業的美玉學姊、已在廈門大學執教的胤藏、和目前正在奮鬥中的憶澤、家瑋、巧瑩、啟裕、和秉軒。感謝 fMRI 組內提供很多幫助和建議的筠安、依寧和艾伶。謝謝 MRI 實驗室的大總管謝昭賢博士及孟錡、威而、永豪、柏暉、子豪、婉婷等學弟妹。在博士班初期提供許多幫助的蘇家豪博士、韋廷，張富雄老師實驗室的新元。感謝你們在我求學階段的幫助與陪伴。

最後，感謝一直支持我的家人，已過世的爺爺張東清、父親張森源、母親李玲淑、弟弟哲榕，一直不離不棄的內人伊慧。感謝你們一路的忍耐、支持與鼓勵，我才能有機會完成這些研究，謝謝你們！

張哲維 謹識

2014 年 1 月於台大

## 中文摘要



本研究的目的是在於利用二維特徵法從大腦磁振造影影像中擷取資訊。在現今大數據的時代裡，我們已經有足夠的技術和資源去收集分析世界上所有醫院及研究單位裡的大腦影像；且在多方努力之下，在不久的將來或將成為事實。因此，迫切需要簡單而有效的方法，可以從各種主要的大腦磁振造影影像中擷取資訊，應用於機器學習的演算法上。

本研究利用二維特徵法從三種重要的大腦磁振造影影像中擷取資訊。首先，我們利用區域二維特徵描述大腦磁振造影解剖影像的形態，並將此特徵用來訓練支持向量機模型，使之能夠利用大腦解剖影像分類注意力不足過動症與正常的孩童；結果顯示，利用此法可以達到 0.6995 的正確率。再者，傳統上比較解剖影像的特徵時，必須先將原始大腦影像轉換至標準化的大腦圖譜上，在相同的圖譜座標系中方能進行比較；在此，區域二維特徵可以直接從未轉換的原始影像上擷取資訊，避免掉一些不必要的轉換和可能引入的雜訊；我們將此法應用在大腦磁振造影解剖影像及擴散磁振造影影像上，並藉此訓練可依大腦結構預測受測者年齡的支持向量機模型；此模型平均絕對誤差最佳可至 5.62 歲。最後，我們試著把相同的概念套用在功能性磁振造影上，利用自行設計的二維特徵法，藉以描述靜息狀態功能性磁振造影所產生的資料，並據以從正常人中偵測注意力不足過動症與精神分裂症之患者；依此法所學習的模型，其正確率較直接使用傳統分析方法為高，在區分精神分裂症患者與正常控制組上可達 0.78，在區分注意力不足過動症患者與正常控制組上可達 0.628。

實驗結果顯示，無論是解剖影像、擴散磁振造影影像、或功能性磁振造影影像，皆可使用二維特徵法擷取其中的資訊。由於其簡單且有效的特性，此法相當適合用於未來大規模的大腦科學相關之實驗及研究。



關鍵字：磁振造影、擴散磁振造影、功能性磁振造影、靜息狀態功能性磁振造影、  
機器學習、模式辨識、注意力不足過動症、精神分裂症



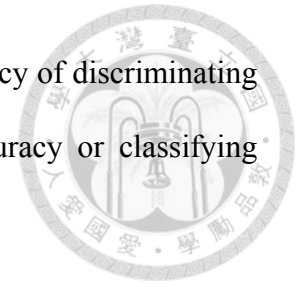
# ABSTRACT



This study aimed to build binary methods to extract efficient information from structural brain magnetic resonance (MR) images and functional brain activities. In the era of big data, to collect and analyze all the brain images in hospitals all over the world is technologically possible and might be achieved in the near future. Therefore, simple and effective methods for machine learning algorithms to extract sufficient information from various brain MR images to build classification or regression models based on numerous brain images are critical.

In this study, we used binary methods to extract information from three different types of brain MR images. First, we implemented local binary patterns (LBP) to describe anatomical brain morphology and used those patterns to train support vector machine models to classify the attention deficit-hyperactivity disorder (ADHD) subjects from normal ones. As a result, the best accuracy we achieved was 0.6995. Second, different from the traditional methods, which all brain images should be normalized to a standard template to be compared in same atlas coordinates, the LBP was used to extract information from unnormalized brain anatomical images and diffusion tensor imaging. We then constructed age estimation models by that extracted information to show the discriminative power of this approach. The best test result mean absolute error of that model equals 5.62 years. Third, following the same line of thought, a binary mapping method was designed and introduced to detect schizophrenia and ADHD patients using resting-state functional MRI data. Compared with traditional cross-correlation network analysis, proposed models exhibits better performance in detecting

schizophrenia and ADHD. Based on our results, the best test accuracy of discriminating schizophrenia from normal subjects was 0.78. The best test accuracy or classifying ADHD from control subjects was 0.628.



Results showed those simple binary methods are useful for extract information from structural and functional brain MR images. Those methods are good candidates to be used in large-scale brain science or medicine related researches.

Key words: MRI, functional MRI, machine learning, big data, local binary pattern, brain age estimation, pattern recognition, attention deficit hyperactivity disorder, ADHD, schizophrenia

# CONTENTS



口試委員會審定書 .....	i
誌謝 .....	ii
中文摘要 .....	iii
ABSTRACT .....	v
CONTENTS .....	vii
LIST OF FIGURES .....	xv
LIST OF TABLES .....	xix
<b>Chapter 1 Introduction.....</b>	<b>1</b>
1.1 Background.....	1
1.1.1 Brain MR Images .....	2



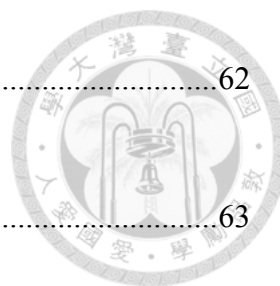
1.1.2	Data Driven Method and Machine Learning Methods	5
1.1.3	Brain MR Images and Machine Learning Methods	6
1.1.4	Neuroimaging Data-sharing Initiative	7
1.2	Motivation and Purpose	9
1.3	Voxel Based Morphology (VBM)	11
1.4	Texture Analysis	14
1.5	The structure of the dissertation	14
<b>Chapter 2</b>	<b>Local Binary Patterns (LBP)</b>	<b>16</b>
2.1	The Development and History of the LBP	16
2.2	LBP of 2D Brain MR Images	19
2.2.1	Local Binary Patterns (LBP)	19
2.2.2	Applications in Medical Images and Brain MR Images	22
2.3	Uniform Patterns	23



2.4	LBP of 3D Brain MR Volumes .....	26
2.4.1	Spatiotemporal LBP - LBP on Three Orthogonal Planes (LBP-TOP) .....	26
2.4.2	Applications in Medical Images and Brain MR Images .....	28
2.5	Support Vector Machine (SVM) .....	28
2.6	Framework of Analysis Using LBP and SVM .....	29
2.6.1	Build Classification Models Using SVM and LBP-TOP .....	29
2.6.2	Feature Selection and Building Regression Models Using SVR and LBP .....	32
<b>Chapter 3</b>	<b>Properties of LBP-TOP .....</b>	<b>36</b>
3.1	Materials and Methods .....	37
3.1.1	Participants .....	37
3.1.2	Parcellations .....	37
3.1.3	Evaluation .....	38



3.2	Parameters of LBP .....	39
3.3	The Efficiency of Uniform Patterns .....	42
3.4	The Order of LBP Coding .....	43
3.5	The Effects of Brain MR Image SNR.....	45
3.6	The Effects of Brain MR Image Resolution.....	48
<b>Chapter 4 Age Estimation Using Unnormalized MR Brain Images .....</b>		<b>50</b>
4.1	Introduction .....	50
4.2	Materials and Methods .....	53
4.2.1	Subjects and Data Acquisition .....	53
4.2.2	Data preprocessing.....	55
4.2.3	Extracting the LBP-TOP Histogram .....	55
4.2.4	Atlas Registration.....	58
4.2.5	Support Vector Regression and Feature Ranking.....	60



4.2.6	Evaluations .....	62
4.3	Results and Discussion .....	63
4.3.1	Age estimation using T1WI .....	63
4.3.2	Age estimation using DTI data .....	70
4.3.3	Atlas Registration .....	75
4.3.4	Learned Model .....	76
4.3.5	Brain Maturation and Aging .....	77
4.4	Conclusion .....	80
<b>Chapter 5</b>	<b>ADHD classification using Local Binary Patterns .....</b>	<b>82</b>
5.1	Introduction .....	82
5.2	Materials and Methods .....	85
5.2.1	Participants .....	85
5.2.2	Diagnostics of ADHD .....	86



5.2.3	Data preprocessing .....	89
5.3	Results .....	99
5.3.1	LBP-TOP.....	99
5.3.2	Permutation test of basic models.....	107
5.3.3	Feature selection.....	108
5.3.4	Resolutions of brain images .....	110
5.3.5	Tissue types .....	112
5.4	Discussion.....	113
5.4.1	Robust to registration method .....	115
5.4.2	Global effects of ADHD?.....	116
5.4.3	Combining models using feature selection .....	117
5.4.4	Most discriminative tissue.....	117
5.5	Conclusion .....	118



5.6	Supplements.....	119
<b>Chapter 6</b>	<b>Discussion, Conclusion, and Future Works.....</b>	<b>121</b>
6.1	Discussion.....	121
6.1.1	Structural MRI .....	121
6.1.2	Different from Traditional Approaches.....	122
6.1.3	Advantages of Using Binary Patterns and Machine Learning Approaches.....	124
6.1.4	Knowledge Discovery .....	124
6.1.5	Limitations .....	125
6.2	Future Works.....	127
6.2.1	Normal Ranges as Image Biomarkers of Brain Images .....	127
6.2.2	Combine Information for Multivariate Approaches.....	127
6.2.3	Detect ADHD and Schizophrenia Using Functional Connectivity Binary Patterns .....	128



6.3 Conclusion.....	165
REFERENCE .....	166



# LIST OF FIGURES



Figure 1-1 Growth patterns in the use of MRI in selected OECD countries.....	3
Figure 1-2 Flow diagram of the preprocessing steps in standard (left) and optimized (right) VBM.....	12
Figure 2-1 LBP in the field of texture analysis operators.....	17
Figure 2-2 Computation of local binary pattern (LBP) from a 2D image.....	21
Figure 2-3 Examples of some texture patterns encoded by LBP histogram.....	22
Figure 2-4 Uniform Patterns.....	24
Figure 2-5 The 58 different uniform patterns of eight neighbors.....	25
Figure 2-6 Computation of LBP on three orthogonal planes (LBP-TOP) from a 3D volume.....	27
Figure 2-7 the flow diagram of the building and evaluation of a classification using SVM and LBP.....	30

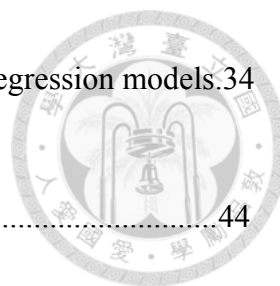


Figure 2-8 the flow diagram of feature selection in classification or regression models.34

Figure 3-1 LBP map using different coding series.....44

Figure 3-2 brain MR images and the corresponding LBP map in different SNR levels. 46

Figure 4-1 Age distribution of the 204 subjects that participated in the study.....54

Figure 4-2 Flow chart for extracting the LBP-TOP histogram from unnormalized MR brain images.....57

Figure 4-3 Example LBP-TOP maps of each tested MR brain image data in this study.58

Figure 4-4 Example of feature selection using linear SVR with whole-brain or GM segment data.....62

Figure 4-5 Scatter plots of age estimation of all 204 subjects using GM segments.....69

Figure 4-6 Scatter plots of age estimation of all 204 subjects using DTI measurements, FA, MD, L1(first eigenvalue), and V1(first eigenvector).....74

Figure 5-1 Overview of the data analysis procedure.....90

Figure 5-2 Examples of brain probability maps based on gray matter (GM), white matter (WM), and CSF.....96

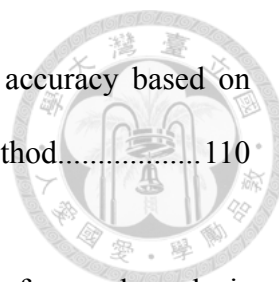


Figure 5-3 Feature selection results of ADHD-TDC classification accuracy based on AAL parcellation and the ART non-linear registration method..... 110

Figure 6-1 Simplify connectivity using resting-state networks referenced analysis (RSNRA). .....131

Figure 6-2 The distribution information not included in the network cross-correlation analysis. ....132

Figure 6-3 The construction of local binary patterns (LBP) and functional connectivity binary patterns (FCBP). .....134

Figure 6-4 Flow chart of network cross-correlation analysis (NCCA), resting-state networks referenced analysis (RSNRA), and functional connectivity binary patterns (FCBP). .....138

Figure 6-5 Connectivity of a single voxel described by ten resting-state networks (RSNs). .....141

Figure 6-6 Illustration of threshold functional connectivity binary patterns (TFCBP). 144

Figure 6-7 Illustration of ordered functional connectivity binary patterns (OFCBP)...146

Figure 6-8 Top 100 features in the classification models of schizophrenia and ADHD using threshold functional connectivity binary patterns with 0.3 CC

thresholds and combining positive and negative features. ....161



# LIST OF TABLES



Table 3-1 the sex classification accuracy of models based on uniform LBP-TOP features with different registration methods, parcellation, and radius of LBP-TOP, using linear-SVM classifiers.....	40
Table 3-2 Sensitivity, specificity, and areas under the ROC curve (AUC) of Table 3-1 the sex classification accuracy of models based on uniform LBP-TOP features with different registration methods, parcellation, and radius of LBP-TOP, using linear-SVM classifiers.....	41
Table 3-3 the sex classification accuracy of models based on LBP-TOP features with and without converting to uniform patterns. ....	43
Table 3-4 the sex classification accuracy of models based on LBP-TOP features with different encode methods, 1mm resolution, 2mm radius, non-linear registration, and uniform patterns.....	44
Table 3-5 the sex classification accuracy of models based on LBP-TOP features with different SNR level, 1mm resolution, 2mm radius, non-linear registration, and uniform patterns.....	46



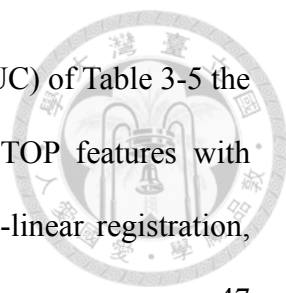


Table 3-6 Sensitivity, specificity, and areas under the ROC curve (AUC) of Table 3-5 the sex classification accuracy of models based on LBP-TOP features with different SNR level, 1mm resolution, 2mm radius, non-linear registration, and uniform patterns. ....47

Table 3-7 the sex classification accuracy of models based on LBP-TOP features with different resolution, radius, registration method, uniform patterns, and combined all brain regions. ....49

Table 4-1 Mean absolute error (MAE) of brain age estimation by LBP-TOP and RBF SVR using T1WI data and 2048 features in various brain regions. ....64

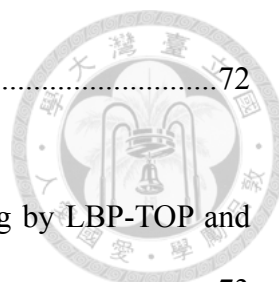
Table 4-2 Mean absolute error (MAE) of brain age estimation by LBP-TOP and RBF SVR using T1WI data in various brain regions.....66

Table 4-3 Mean absolute error (MAE) of brain age estimation by LBP-TOP and RBF SVR using T1WI data and 2048 features in combined brain regions. ....67

Table 4-4 Mean absolute error (MAE) of brain age estimation by LBP-TOP and RBF SVR using T1WI data in combin.....68

Table 4-5 Mean absolute error (MAE) of brain age estimation by LBP-TOP and RBF SVR using DTI data and 2048 features in combined brain regions. ....72

Table 4-6 Mean absolute error (MAE) of brain age estimation by LBP-TOP and RBF



SVR using DTI data in combined brain regions.....72

Table 4-7 Mean absolute error (MAE) of brain age estimation using by LBP-TOP and RBF SVR DTI data in various brain regions.....73

Table 4-8 The proportion of features from specific brain regions used in the top 256 features of the age estimation models.....76

Table 4-9 Comparison of age estimation studies using MR brain images. ....77

Table 4-10 Mean absolute error (MAE) of different age ranges of brain age estimation by LBP-TOP and RBF SVR using GM and DTI data (the results of specific subject are plotted in Figure 4-5 and Figure 4-6). ....78

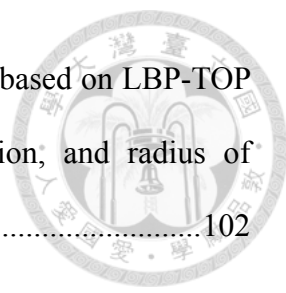
Table 5-1 A brief summary of the different diagnostic criteria used by each site. ....87

Table 5-2 Summary of the ADHD and TDC subjects used in this study.....88

Table 5-3 The age distribution of subjects used in this study. ....88

Table 5-4 (a, b) The ADHD-TDC classification accuracy of models based on LBP-TOP features with different registration methods, parcellation, and radius of LBP-TOP, using 1NN and linear-SVM classifiers alternatively.....101

Table 5-5 Sensitivity, specificity, and areas under the ROC curve (AUC) of (a) in Table



5-4. The ADHD-TDC classification accuracy of models based on LBP-TOP features with different registration methods, parcellation, and radius of LBP-TOP, using 1NN. .... 102

Table 5-6 Sensitivity, specificity, and areas under the ROC curve (AUC) of (a) in Table 5-4. The ADHD-TDC classification accuracy of models based on LBP-TOP features with different registration methods, parcellation, and radius of LBP-TOP, using linear-SVM classifiers. .... 103

Table 5-7 *p*-values of McNemar’s test of 1NN models and linear-SVM models based on LBP-TOP features between different radii (R1, R2, R3). .... 104

Table 5-8 *p*-values of McNemar’s test of 1NN models and linear-SVM models based on LBP-TOP features between different registrations. .... 105

Table 5-9 *p*-values of McNemar’s test of linear-SVM models based on rs-fMRI data and LBP-TOP features using non-linear registration. .... 106

Table 5-10 Permutation test of some results in Table 5-4 ..... 107

Table 5-11 ADHD-TDC classification results using LBP-TOP based on the ART non-linear registration method. .... 109

Table 5-12 Feature selection results of ADHD-TDC classification accuracy of different resolution of source images. .... 111

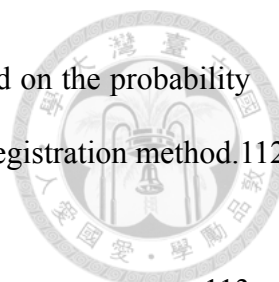


Table 5-13 The ADHD-TDC classification accuracy of models based on the probability map of different brain tissues using the ART non-linear registration method.112

Table 5-14 *p*-values of McNemar’s test of results in Table 5-12..... 113

Table 6-1 Cross-validation results of resting-state networks referenced analysis (RSNRA) and network cross-correlation analysis (NCCA).....149

Table 6-2 *p*-values of McNemar’s test of results in Table 6-1..... 150

Table 6-3 Cross-validation results of threshold functional connectivity binary patterns (TFCBP). ..... 152

Table 6-4 Cross-validation results of threshold functional connectivity binary patterns (TFCBP). ..... 153

Table 6-5 *p*-values of McNemar’s test of results in Table 6-3 and Table 6-4..... 154

Table 6-6 Cross-validation results of ordered functional connectivity binary patterns (OFCBP and absolute OFCBP, ABS-OFCBP) and network cross-correlation analysis (NCCA)..... 156

Table 6-7 *p*-values of McNemar’s test of results in Table 6-6..... 157

Supp. Table 5-1 List of all subjects. ....119





# Chapter 1

## Introduction

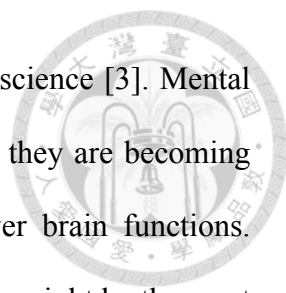
*"Simple models and a lot of data trump more elaborate models based on less data."*

- Peter Norvig,

The unreasonable effectiveness of data [2].

### 1.1 Background

To reveal how the brain works is long to be a great unsolved challenge. The last three

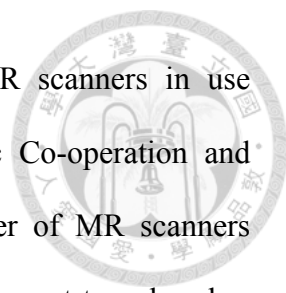


decades have seen growing importance placed on research in neuroscience [3]. Mental disorders are the most important topics in those researches because they are becoming the major threats to our health, and could be the keys to uncover brain functions. Nevertheless, the human brain built by more than 100 billion neurons might be the most complex organ in our body. It is hard to use the change of specific brain region to explain the brain developments or detect many mental disorders. Combining the information from all brain regions might provide another insight to identify the brain maturation, aging, or disorders. In this study, we tried to extract information from brain magnetic resonance (MR) images and construct useful classification and regression models by structural or functional information from all brain regions.

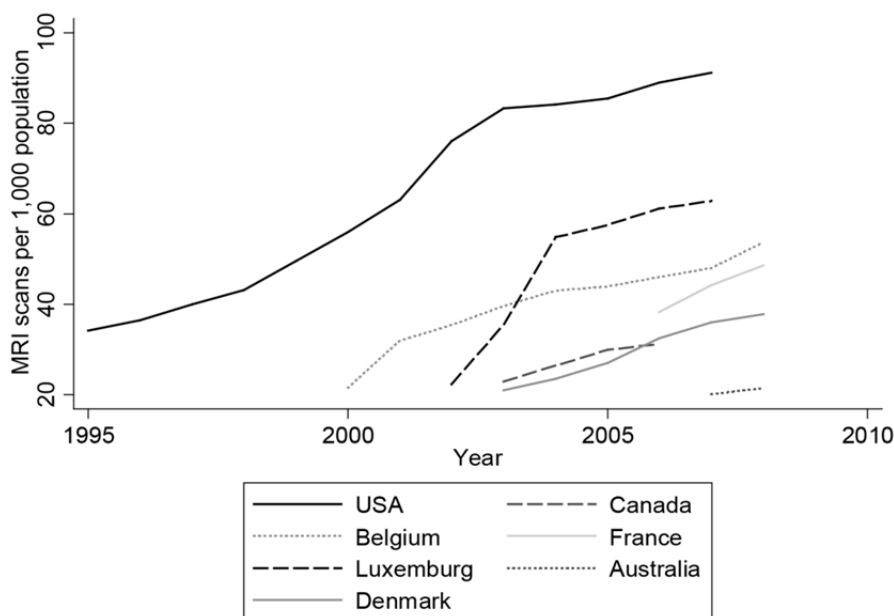
### 1.1.1 *Brain MR Images*

Magnetic resonance imaging (MRI) is a medical imaging technique used to noninvasively investigate the anatomy or function of the human brains and bodies. With strong magnetic fields and radio waves, MRI scanners can build the anatomical details of the body. Because of the ability of noninvasive imaging and producing high resolution results, MRI is widely used in hospitals for medical diagnosis. Besides, MRI can also be used to explore the brain blood oxygen level dependent (BOLD) signal. Therefore, one can investigate brain functional activity using a series of MRI images, which is known as functional MRI (fMRI). FMRI is very helpful in research organizations for analyzing the human brain functions.

In recent decades, brain MR imaging has become an important tool in neurology and



clinical neuroscience [4, 5]. There are approximately 25,000 MR scanners in use worldwide<sup>1</sup>. Based on data from the Organization for Economic Co-operation and Development (OECD), as shown in Figure 1-1, either the number of MR scanners installed or the number of MR scans performed increased rapidly in past two decades. Meanwhile, the MRI technologies are getting faster, better SNR, and higher resolution every day [4]. Soon the cheaper, speedier, and more detailed brain images full of important neurological information would be popular and convenient to get.



**Figure 1-1 Growth patterns in the use of MRI in selected OECD countries.**

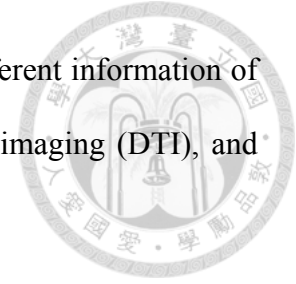
A reprint of the Figure 2 of [6].

---

<sup>1</sup> [http://www.magnetic-resonance.org/MagRes%20Chapters/21\\_02.htm](http://www.magnetic-resonance.org/MagRes%20Chapters/21_02.htm)



There are three main types of brain MR images to provide very different information of the brain, that is, the structural brain images, the diffusion tensor imaging (DTI), and functional MRI.



### Structural MRI

Structural MRI, especially the T1-weighted images (T1WI), provides the details of brain structural information, including the shape, size, and the integrity of gray and white matter. In many brain morphometric studies, the volume, shape, or thickness of gray matter structure have been measured to provide indications of diseases or mental disorders [7].

### Diffusion Tensor Imaging (DTI)

Diffusion tensor imaging (DTI) is the MRI technique that measures the restricted diffusion of water molecules in tissue and offers directional information of neural tracts in the brain [8]. DTI can provide various information of brain microstructure and has proven its value in both brain research as well as clinical applications [9, 10]. With yielding more information rather than structural brain MR images, DTI is also an important type of brain MR images in clinical used.

## Resting-State Functional MRI (rs-fMRI)

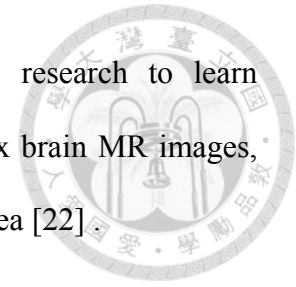


In recent years, the resting-state functional magnetic resonance imaging (rs-fMRI) has become a novel technique for studying the mental illness [11-13]. Different from the traditional task-based or stimulus-based functional MRI, rs-fMRI investigates spontaneous synchronous activations between brain regions occurring in the resting-state. Inferred on the basis of blood-oxygen-level dependence (BOLD) response time-series data, the neural activity revealed by rs-fMRI can be easily obtained from normal subjects or patients [13]. A number of studies have proven the resting-state functional connectivity can detect the differences between patients and controls in various neurological and psychiatric disorders [12, 14, 15]. Promised results have been obtained for various neurological disorders, such as Alzheimer's disease [16-18], and multiple sclerosis [19, 20].

### 1.1.2 *Data Driven Method and Machine Learning Methods*

Traditionally, there are two approaches to learn useful information from data [21]. One is the traditional statistical methods which assume that the data are generated by a given model and be validated by goodness-of-fit tests and residual examination. The other is data-driven approaches, which use the algorithmic models to learn rules from data and be validated by cross-validation and predictive accuracy. For analyzing simple mechanisms, the first approach is good and intuitive, such as the commonly used linear regression or logistic regression. However, while facing complex mechanisms, it is always hard to find a proper model to describe the data. Second approach is more useful

in this situation. Therefore, we used second approach in this research to learn classification or regression models by algorithms from the complex brain MR images, which is called “machine learning“ approach in computer science area [22].



We have seen mounting evidence of the usefulness of mining information from vast data, including detecting influenza epidemics using search engine query data [23], translating language using billions of web pages [22], finding human faces in images [24], recognizing of vehicle license plates in video sequence [25], and so on [22, 26]. Based on the view point of machine learning, the brain MR images stored in the hospital's data centers are treasure-trove providing information to reveal the mystery of brain mechanism, maturation, aging, and mental disorders.

### 1.1.3 *Brain MR Images and Machine Learning Methods*

For years, the brain MR images have been applied to patients with psychiatric or neurological disorders to find possible biomarkers for diagnosis. Using traditional statistical methods, those researches has revealed structural and functional alterations in several disorders, for example, major depression [27], anxiety disorder [28], Alzheimer's disease [29], and schizophrenia [30]. Although those findings might contribute to the understanding of disorders, they have had minimal clinical impact. Neurologists and psychiatrists still relied on traditional diagnostic and prognostic tools.

The neuroimaging studies used traditional statistical methods could show the differences between patients and controls at group level. In contrast, doctors should make clinical decisions about individuals in clinical. In clinical circumstances, knowing

the group differences does not provide too much useful information.



As we mentioned in 1.1.1, the machine learning approaches are good for mining useful information from complex brain MR images data and make inference at the level of the individual. Over the past few years, there has been growing interest within the neuroimaging community in the use of supervised machine learning [31]. However, most of these studies used complex analysis and tested in small datasets, which cannot provide strong and robust inference. In this study, we tried to use simple information extraction method, the LBP-TOP texture analysis and use big database to train strong and robust classification or regression models.

#### 1.1.4 *Neuroimaging Data-sharing Initiative*

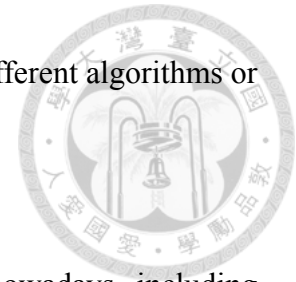
In neuroscience community, there has been an increase in sharing of neuroimaging data in recent years [32]. Many scientific fields have shown the benefits of sharing data. For example, the astronomy<sup>2</sup>, natural history [33], and the most famous, the GenBank and Hapmap in genetics have led to many scientific discoveries [34]. Neuroimaging research is very costly and time-consuming. To share the imaging data and enable to reuse them can reduce the cost of research, provide more reproducible. Moreover, just like the UCI machine learning repository<sup>3</sup> in computer science area, publicly available

---

<sup>2</sup> <http://www.sdss.org/>

<sup>3</sup> <http://archive.ics.uci.edu/ml/>

databases can also be used as the standard test sets for comparing different algorithms or analysis methods.



Hence, more and more neuroimaging data is publicly available nowadays, including Alzheimer's Disease Neuroimaging Initiative (ADNI<sup>4</sup>), the NIH MRI Study of Normal Brain Development<sup>5</sup>, the National Database for Autism Research (NDAR<sup>6</sup>), the Open Access Series of Imaging Studies (OASIS<sup>7</sup>), the International Neuroimaging Data-sharing Initiative (INDI<sup>8</sup>), and increasing research centers trying to make their data public [32].

Now there are roughly thousands of brain MR images with normal or abnormal subjects publicly available on the network. To analyze those huge neuroimaging data, and might be much larger in near future, simple and effective methods to extract smaller but discriminative information are critical.

---

<sup>4</sup> <http://www.adni-info.org/>

<sup>5</sup> [https://nihpd.crbs.ucsd.edu/nihpd/info/data\\_access.html](https://nihpd.crbs.ucsd.edu/nihpd/info/data_access.html)

<sup>6</sup> <http://ndar.nih.gov/>

<sup>7</sup> <http://www.oasis-brains.org/>

<sup>8</sup> [http://fcon\\_1000.projects.nitrc.org/](http://fcon_1000.projects.nitrc.org/)



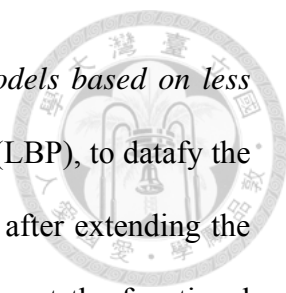
## 1.2 Motivation and Purpose

This study aimed to build simple binary methods to extract efficient information from structural brain MR images and rs-fMRI. And then we evaluated that information by constructing discriminated classification and regression models.

Brain MR image is high-dimension data showing the complex brain structure. While the MR machine producing more images of higher resolution, more information about the brain structure would be provided. Theoretically, it could be more easily for neurologists to find out the abnormal area. Unfortunately, the time of each neurologist is limited. Within the limited time, it is harder to check every detail structure within images with higher resolution and more slices [35]. Therefore, it is very helpful if an algorithm can automatically analyze, classify, or highlight the abnormal brain images. The algorithm could screen the image data and classify mental disorder from normal subjects based on the brain images stored in the hospital's data center or images publicly available in the Internet. Thus, the algorithm cannot only help neurologist to make better diagnosis but also can be used as a discovery science tool [36].

To investigate brain functions or disorders from vast of brain MR images, machine learning algorithms should effectively recognize and use the information embedded in those images. The way how to convert the original digital images to informations known by algorithms is one of the most critical steps in machine learning process.

Following the experience of Google's artificial-intelligence guru, Peter Norvig, who



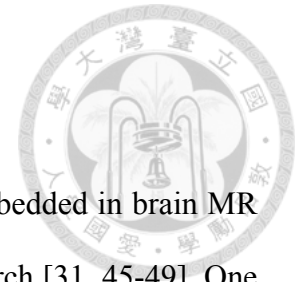
wrote, “*Simple models and a lot of data trump more elaborate models based on less data.*”[2], we used simple texture analysis, the local binary patterns (LBP), to datafy the structural brain MR images, such as the T1WI and DTI. Moreover, after extending the idea of structural binary patterns, we also try to design and implement the functional binary patterns to extract information from rs-fMRI data.

The purpose of this study is to construct simple and effective methods to extract information from structural or functional brain MR images using binary patterns and build discriminative models by machine learning algorithms.

Three main different types of brain MR images are used to verify the approaches. For structural images, brain T1WI and DTI images would be used. For functional brain images, the rs-fMRI data would be used.

To test the performance of introduced information extraction methods, we used several brain MR image databases. Two types of images were tested. One is those brain images with obvious visible structural changes. Many studies have shown cerebral changes during brain maturation and aging [37-44]. Therefore, we use the binary patterns to estimate age based on brain MR images in this type. Another type is those brain images without obvious structural difference, such as many mental disorders. In this type, Attention-Deficit/Hyperactivity Disorder (ADHD) and schizophrenia are used to test our approaches.

### 1.3 Voxel Based Morphology (VBM)

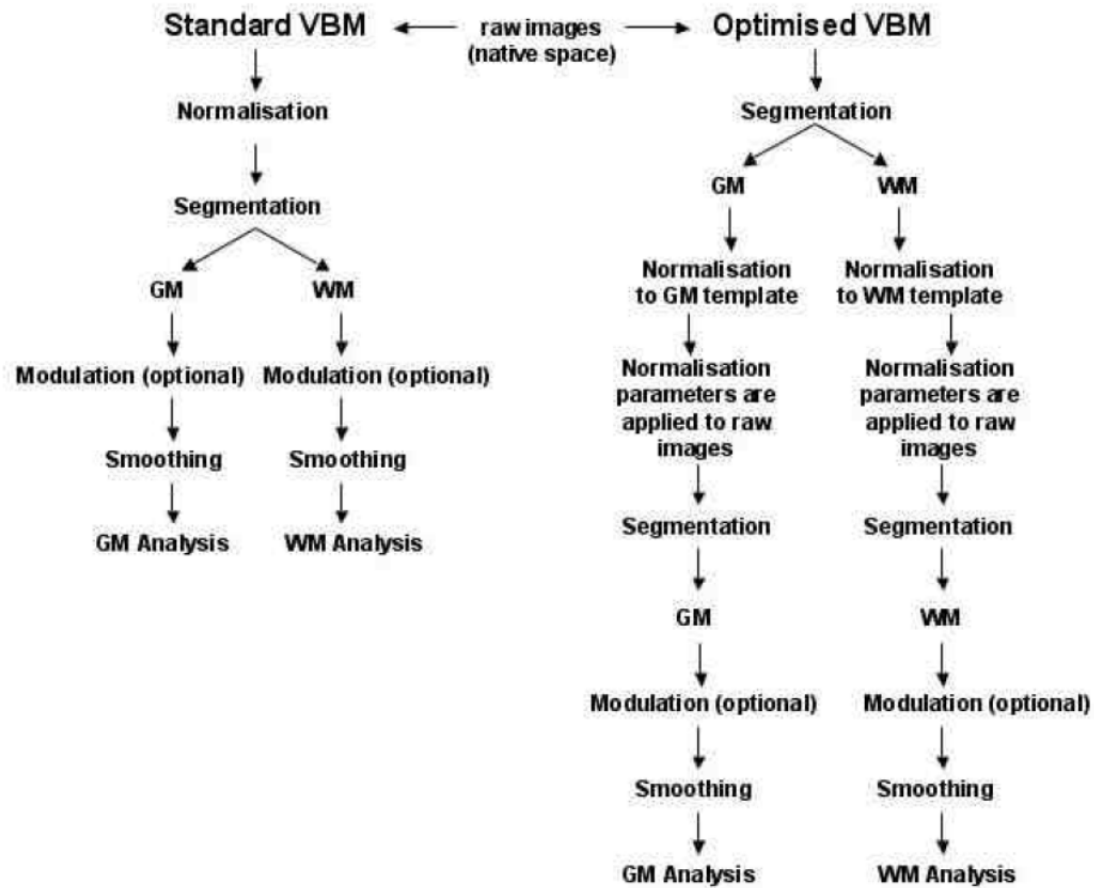


Many researchers have attempted to reveal imaging biomarkers embedded in brain MR images to facilitate clinical diagnosis or enhance neurological research [31, 45-49]. One of the major goals of these efforts is to build useful classification or regression models using machine learning approaches to classify or evaluate various neurological diseases [31]. As the old computer saying, "garbage in, garbage out", before structuring a good model, we should first extract effective features from MR brain images. Therefore, an automatic, robust and efficient method for extracting useful information from brain morphology is very important.

Most of the existing studies used voxel based morphology (VBM) approaches [50] to analyze the different brain structure. VBM, usually based on high-resolution T1WI, is the comparison of local gray matter concentration at every voxel between groups. To conduct VBM, one should segment the brain T1WI into tissue classes, spatially normalized, smoothed, and analyzed the resulted data voxel-by-voxel in general linear or nonparametric models (Figure 1-2). This technique started in the mid 1990's [51] and became a widely applied method in computational neurosciences[52, 53], including: Schizophrenia [51], dementia [54], aging [41], cocaine abuse [55], Parkinson's disease [56], and many other studies (check [53] for more related studies).

However, it is currently recommended that VBM can only be used to compare data collected at a single MRI scanner [57]. Therefore, it is not a good method to be used in big data researches. Moreover, there are several unsolved problems using this approach.

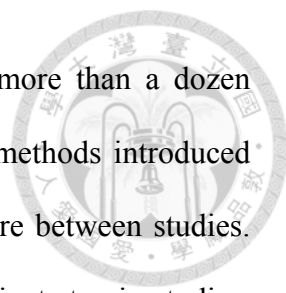




**Figure 1-2** Flow diagram of the preprocessing steps in standard (left) and optimized (right) VBM.

GM = gray matter images; WM = white matter images. This is a reprint of the Fig. 2.4 of [50]

First, all the subject brain images need to be registered to some standard brain template. Imperfect registration might modify fine structural details of the brain and introduce unexpected bias [58, 59]. Consequently, the results of the analysis could depend on the different registration methods used [49, 58, 60]. Second, there is no standard way to perform registration. Due to the complexity of brain structure, linear affine



transformation is not sufficient for many approaches. Therefore, more than a dozen non-linear registration methods have been introduced [60]. These methods introduced distinct parameters and made it more difficult to repeat and compare between studies. Third, the non-linear registration method is usually the most inefficient step in studies. Fourth and even more troubling, Uylings et al. have shown that the huge individual variability that exists in brain structure makes it nearly impossible to register distinct brains to a standard template using only morphological information [61].

Therefore, if we can extract morphological information from unnormalized source brain images, we cannot only avoid the risk of crashes of several preprocessing steps, but we can also access features without unexpected modification. In Chapter 4, we use a simple texture based feature extraction method based on local binary pattern on three orthogonal planes (LBP-TOP) to extract effective features from unnormalized source brain volumes.

Second, traditional VBM only considered the gray matter (GM) segmentation (or known as GM concentration) as the information source. The process of segmenting brain images to GM, white matter (WM), and cerebrospinal fluid (CSF) could be another step might fail or introduce noise to the image data. In Chapter 4, we also tested both the unmodified whole-brain images and the GM segments as the information source in this work.



## 1.4 Texture Analysis

Texture-based analysis has recently proven to have an excellent power of discrimination and simple implementation in many applications in the computer vision domain [62]. Moreover, in the domain of MR medical images, this method has shown its potential in extracting discriminative information about brain tumors [63, 64], epilepsy [65], Alzheimer's disease [66], and multiple sclerosis [67, 68]. In addition, texture analysis can also provide effective features for tissue characterization and dynamic contrast-enhanced MRI [69-71]. However, most texture analyses of MR images focus on a specific region of 2D brain images. Few studies have used these methods as a discovery science tool to extract useful information from whole-brain data.

## 1.5 The structure of the dissertation

The background of our study was illustrated in this chapter. Chapter 2 will introduce the local binary patterns (LBP) method and the framework of analysis in this study. Moreover, the basic properties of LBP would be tested in Chapter 3. In Chapter 4 and Chapter 5, we used local binary patterns (LBP) to extract useful information from MR brain morphology. First, we found the LBP is robust to brain registration method and can be performed well with the unnormalized brain images. Therefore, we used LBP to extract structural information from unnormalized T1-weighted brain images and diffusion tensor imaging (DTI) and constructed age estimation models based on brain T1WI and DTI images in Chapter 4. Then we use registered brain images as the

information source to build ADHD classification models in Chapter 5. Finally, the advantages, limitations and future works of those approaches are discussed in Chapter 6. In 6.2, we extended the idea of binary pattern distribution from structural MRI to resting-state functional MRI. Functional connectivity binary patterns based on ICA results are introduced and used to build classifications of ADHD and Schizophrenia.

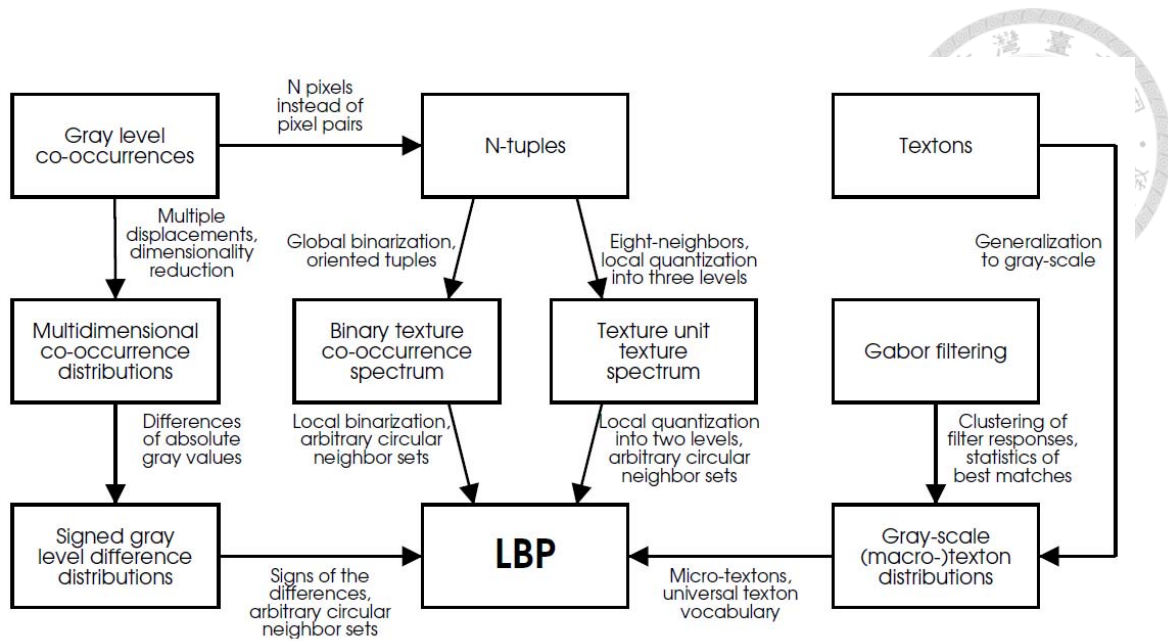


## **Chapter 2**

### **Local Binary Patterns (LBP)**

#### **2.1 The Development and History of the LBP**

The local binary pattern (LBP) is a simple yet very efficient texture operator. LBP encodes the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number. Figure 2-1 shows the relationships of LBP and many well-known texture analysis operators.

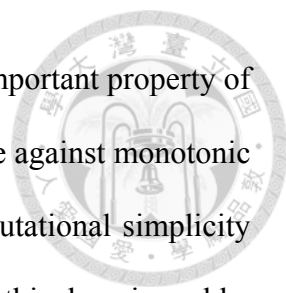


**Figure 2-1 LBP in the field of texture analysis operators.**

A reprint of the Fig. 2. of [72].

The basic LBP was developed by David Harwood in 1992 [73]. The basic idea is that two-dimensional textures can be simply described by pattern and contrast. LBP can separate the pattern information unaffected by monotonic gray scale form contrast. At that time, using feature distributions to classify texture was not very popular ( Using the statistical approach to classify images was later introduced around 2009 [74] ). LBP was first published in 1994 [75] and an extended version with it in 1996 [76]. Then LBP were later used for unsupervised texture segmentation [77]. The results showed better performance than the state-of-the-art approaches and revealed the potential of LBP. After that, the rotation-invariant LBP was introduced in the late 1990s [78].

Around 2000, the theoretical basis of LBP and advanced version of rotation-invariant multiscale LBP operator was published[79, 80], which also introduced the idea of



“uniform patterns” to reduce the feature size of the LBP. The most important property of the LBP operator in 2D computer vision applications is its invariance against monotonic gray level changes. In addition, its discriminative power and computational simplicity let the LBP texture operator has become a very popular approach in this domain and be used in many applications.

In 2004, the LBP features were used as a novel facial representation for face recognition [81]. The LBP features were extracted from different regions of human faces to build as an enhanced feature vector to be used as a face descriptor. This approach was another excellent success of the LBP and has been adopted and further developed by many research groups and companies. This approach has been used to perform face detection, face recognition, gender classification and age estimation based on 2D face images.

The LBP was also used in the motion analysis. It began with the development of a texture based method for modeling the background and detecting moving objects [82]. Then in 2007, the spatiotemporal LBP approaches were proposed [83]. Those approaches soon began the basis for motion and activity analysis and has been used for recognizing facial expression [83], face and gender recognition from videos [84], and human activity recognition [85-87]. Recently, the LBP was used in other applications outside the computer vision domain; we will discuss the applications in medical images in 2.2.2.



## 2.2 LBP of 2D Brain MR Images

### 2.2.1 Local Binary Patterns (LBP)

Texture analysis based on local binary patterns (LBP) has recently been shown to have excellent discriminative power for many applications in the domain of computer vision [73, 80]. LBP was originally designed to extract features from various textured images, such as organic fibers, wood, and fabric [76]. After decades of development, it was also found to be useful for extracting the features from other types of images, such as face description [88], image segmentation, and other applications [73]. Furthermore, it can be used as a spatiotemporal descriptor for motion and activity analysis [83]. In the domain of computer vision, LBP is an efficient and robust method for extracting information from morphology [73].

LBP is a simple and efficient image texture operator introduced by Ojala et al. [76, 80]. Figure 2-2 shows the three steps for computing LBP on 2D images. The  $LBP_{P,R}$  operator can be defined as

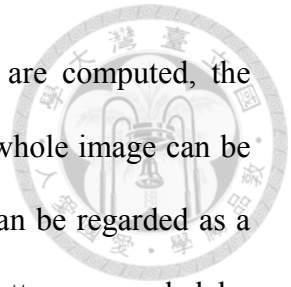
$$LBP_{P,R} = \sum_{p=0}^{P-1} \text{sign}(v_p - v_c) 2^p \quad (2-1)$$

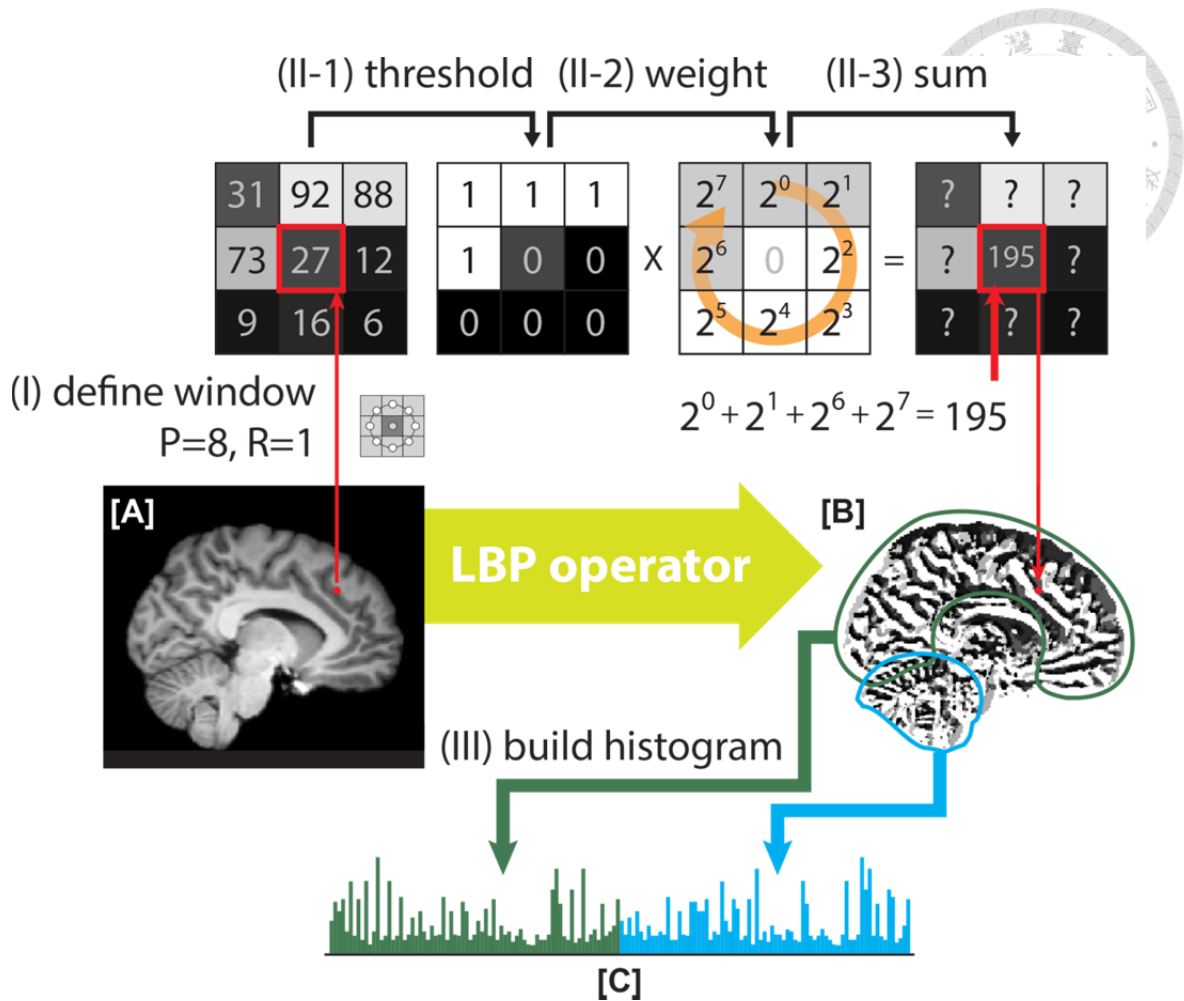
$$\text{sign}(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

where  $v_c$  and  $v_p$  are the values of the center pixel and neighborhood pixels with radius  $R$ , respectively,  $P$  is the total number of neighborhood pixels, and  $R$  is the



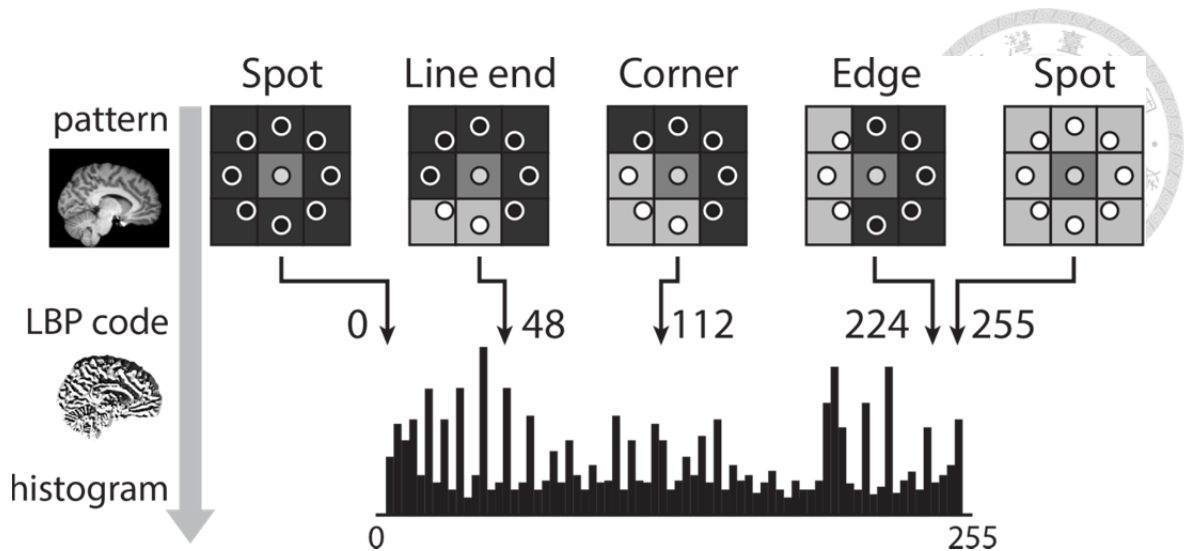
radius in pixel. After the LBP codes for all voxels in an image are computed, the histogram of the codes computed over specific regions or over the whole image can be used as a texture descriptor. Therefore, each bin of the histogram can be regarded as a "micro-texton" encoded by LBP [89]. Figure 2-3 demonstrates patterns encoded by these histogram bins. Any morphological changes would modify the distribution of the codes, resulting in alterations to the histogram. Therefore, the histogram of the computed LBP codes is a good descriptor for comparing changes between images.





**Figure 2-2 Computation of local binary pattern (LBP) from a 2D image.**

LBP define a mapping of all pixels from the grayscale pattern space [A] to the binary pattern space [B] by three simple steps. **Step (I)**: Define a small window by a radius R and number of neighborhood pixels P. **Step (II-1)**: Threshold the neighborhood of each pixel by the value of the center pixel and consider the result as a binary number. **Step (II-2, II-3)**: The code of the center pixel is given as a weighted sum of its thresholded neighboring pixels. **Step (III)**: After the LBP of an image was computed, the combined histogram [C] over regions is used as the texture feature.



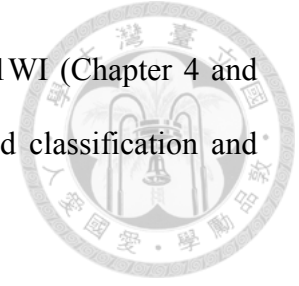
**Figure 2-3** Examples of some texture patterns encoded by LBP histogram.

### 2.2.2 Applications in Medical Images and Brain MR Images

As we have shown in 2.1, after 2000, the LBP became a popular method in computer vision domain and be used in many applications. Some medical applications also use LBP to classify different medical images. Rosdi et al. used LBP to perform the finger vein recognition [90]. Unay et al. and Tommasi et al. used LBP to construct the medical image search and retrieval system [91, 92].

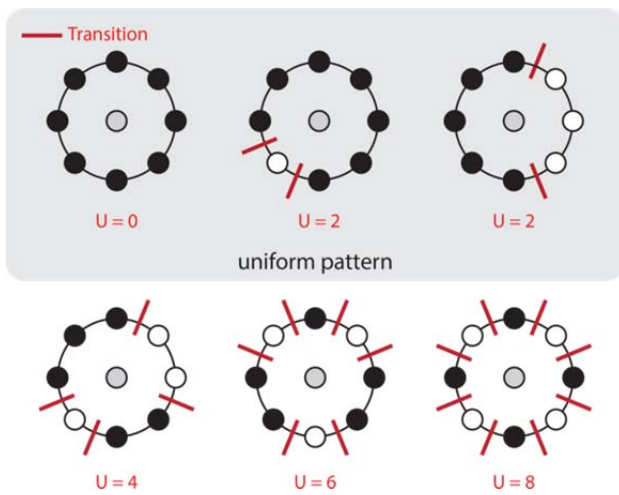
As for the brain MR images, the important property of the LBP's invariance against monotonic gray level changes makes it's a good candidate for comparing MR images collected from different research sites. Unay et al. have shown the robustness of LBP to the intensity inhomogeneity of 2D brain MR images. They also show the texture-based method is better than intensity-based method for image search and retrieval [92-94].

Therefore, we use LBP-TOP to extract useful information from T1WI (Chapter 4 and Chapter 5) and diffusion tensor imaging (DTI, Chapter 4) to build classification and regression models for patients and normal subjects.



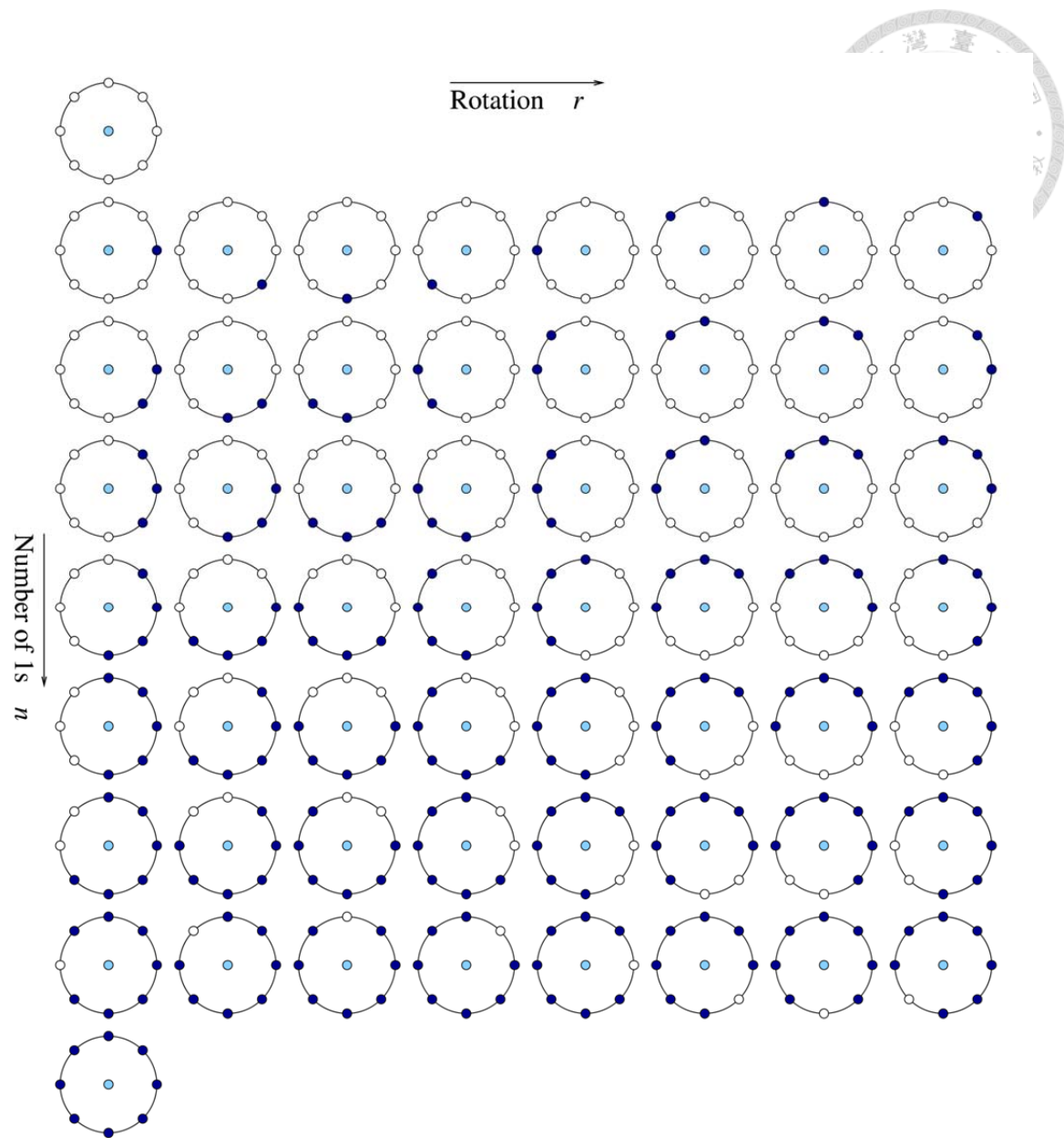
## 2.3 Uniform Patterns

Ojala et al. also provided a simple method to map the histogram bins of LBP to a smaller "uniform patterns" space [80]. They found that non-uniform patterns rarely exist in many image classification applications. Considering the LBP as circular,  $U$  is the number of bitwise transitions from 0 to 1 or vice versa. A LBP pattern is called uniform if its  $U$  is at most 2 (Figure 2-4, Figure 2-5). While mapping to uniform patterns, a separate output label is assigned to each uniform pattern, and all non-uniform patterns are assigned to a single label (usually the code 0, as the smallest LBP code). In the case of LBP with eight neighbors, the length of the histogram bins of one image is reduced from 255 to 59 after mapping to uniform patterns (all of these 59 uniform patterns can be found in [73]). Therefore, the histogram bins of one region of LBP-TOP can be reduced from 765 to 177.



**Figure 2-4 Uniform Patterns.**

While considering the LBP as circular,  $U$  is the number of bitwise transitions from 0 to 1 or vice versa. A LBP pattern is called uniform if its  $U$  is at most 2.



**Figure 2-5 The 58 different uniform patterns of eight neighbors.**

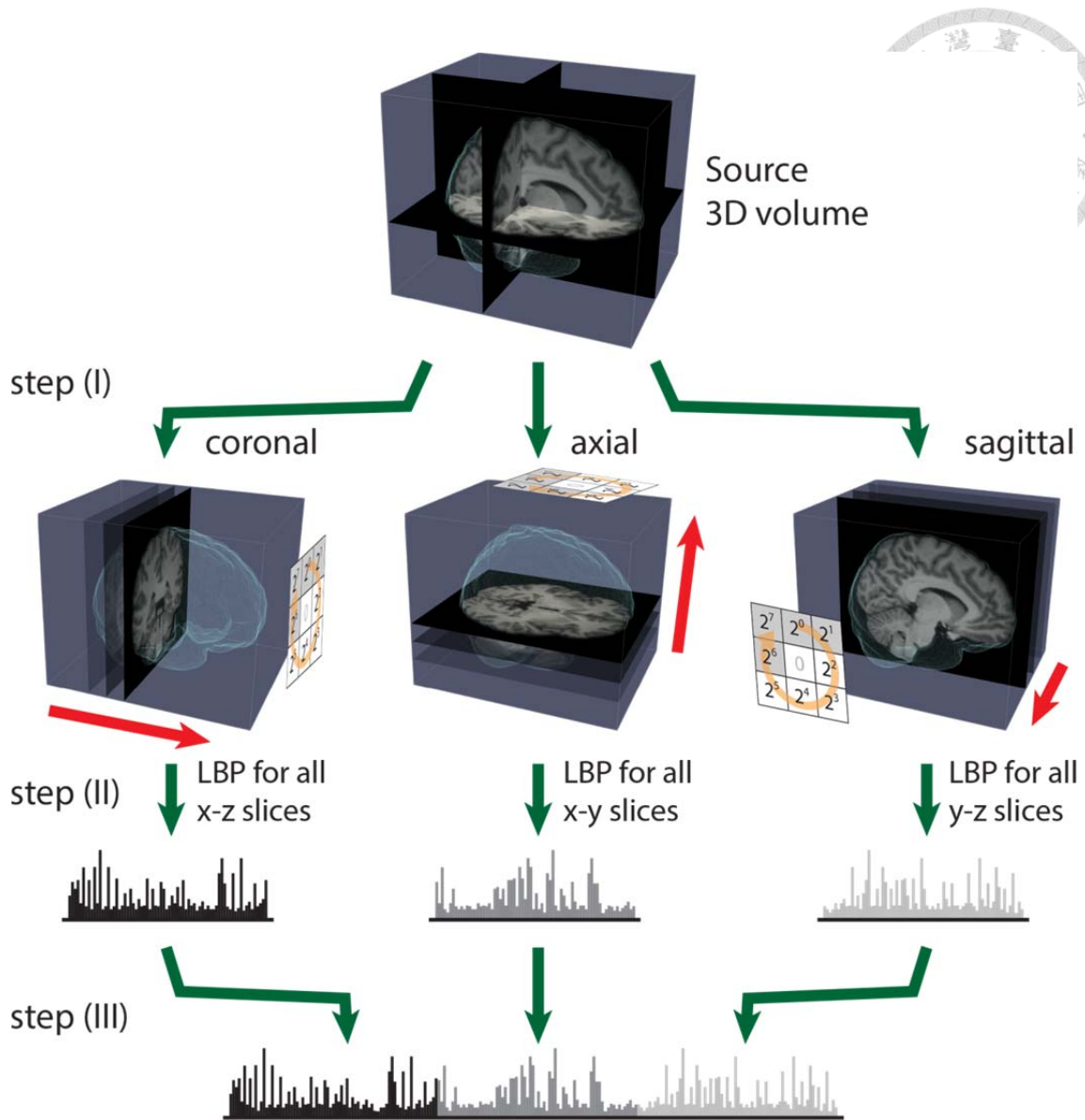
A reprint of the Fig. 2.4 of [73].

## 2.4 LBP of 3D Brain MR Volumes



### 2.4.1 *Spatiotemporal LBP - LBP on Three Orthogonal Planes (LBP-TOP)*

For 3D data, Zhao et al. have proposed simplifying spatiotemporal descriptors by concatenating LBP on three orthogonal planes (LBP-TOP), i.e., the xy, xt and yt planes [83]. Here, we used LBP-TOP to describe brain volume data. Therefore, we replaced the t dimension with the z dimension. We propose using the same radius for x, y, and z for LBP-TOP. Figure 2-6 illustrates the specific steps for computing LBP-TOP on 3D-volume data.



**Figure 2-6 Computation of LBP on three orthogonal planes (LBP-TOP) from a 3D volume.**

**Step (I):** Define a small window by a radius  $R$  and number of neighborhood voxels  $P$ . **Step (II):** LBP codes are computed on three orthogonal directions ( $x$ ,  $y$ , and  $z$ ). Each voxel is encoded based on those three orthogonal planes ( $xy$ ,  $yz$ , and  $xz$ ). After the LBP of each direction was computed, the histogram over a specific volume (whole brain volume in this example) is the texture feature of that direction. **Step (III):** Combine the histograms of those three directions to build the result histogram as the LBP-TOP feature of the volume.



## 2.4.2 Applications in Medical Images and Brain MR Images

Few medical image applications used LBP-TOP. Gao X. and his coworkers build a brain image retrieval system using LBP-TOP [95, 96]. They study several 3D image retrieval methods and find LBP-TOP is the fastest and most accuracy one. Those studies manifest the power of LBP-TOP to extract information from brain MR images.

## 2.5 Support Vector Machine (SVM)

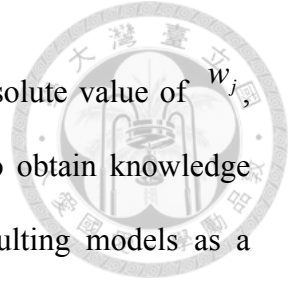
To evaluate the discrimination power of each approach, an efficient and widely used classifier, the support vector machine (SVM), was used [97]. SVM maps training data into high-dimensional feature spaces to adequately separate the hyperplane with a maximal margin [98, 99]. We used linear SVM to evaluate each method used in this work. The widely used LIBLINEAR program was used to implement SVM because of its optimization of linear SVM [100].

Linear SVM is an adequate feature ranking method and has demonstrated as an efficient and useful tool for gene selection, document classification, and many other applications [101-103]. For any test subject  $x$ , the decision function of linear SVM is

$$P(x) = \text{sgn}(\mathbf{w}^T \mathbf{x} + b) \quad (2-2)$$

where  $\mathbf{x}$  is the feature vector,  $b$  is a constant, and  $\mathbf{w}$  is the weight vector. Each

value of  $\mathbf{W}$  denotes the weight of each feature. The larger the absolute value of  $w_j$ , the more important the  $j$ th feature is when deciding the result. To obtain knowledge from the trained linear SVM models, we can use  $\mathbf{W}$  in the resulting models as a relative importance index to determine which features are more useful to discriminate the diseased from the control subjects.



## **2.6 Framework of Analysis Using LBP and SVM**

### *2.6.1 Build Classification Models Using SVM and LBP-TOP*

The flow diagram in Figure 2-7 shows the steps of process of the building and evaluation of classification using SVM and LBP.

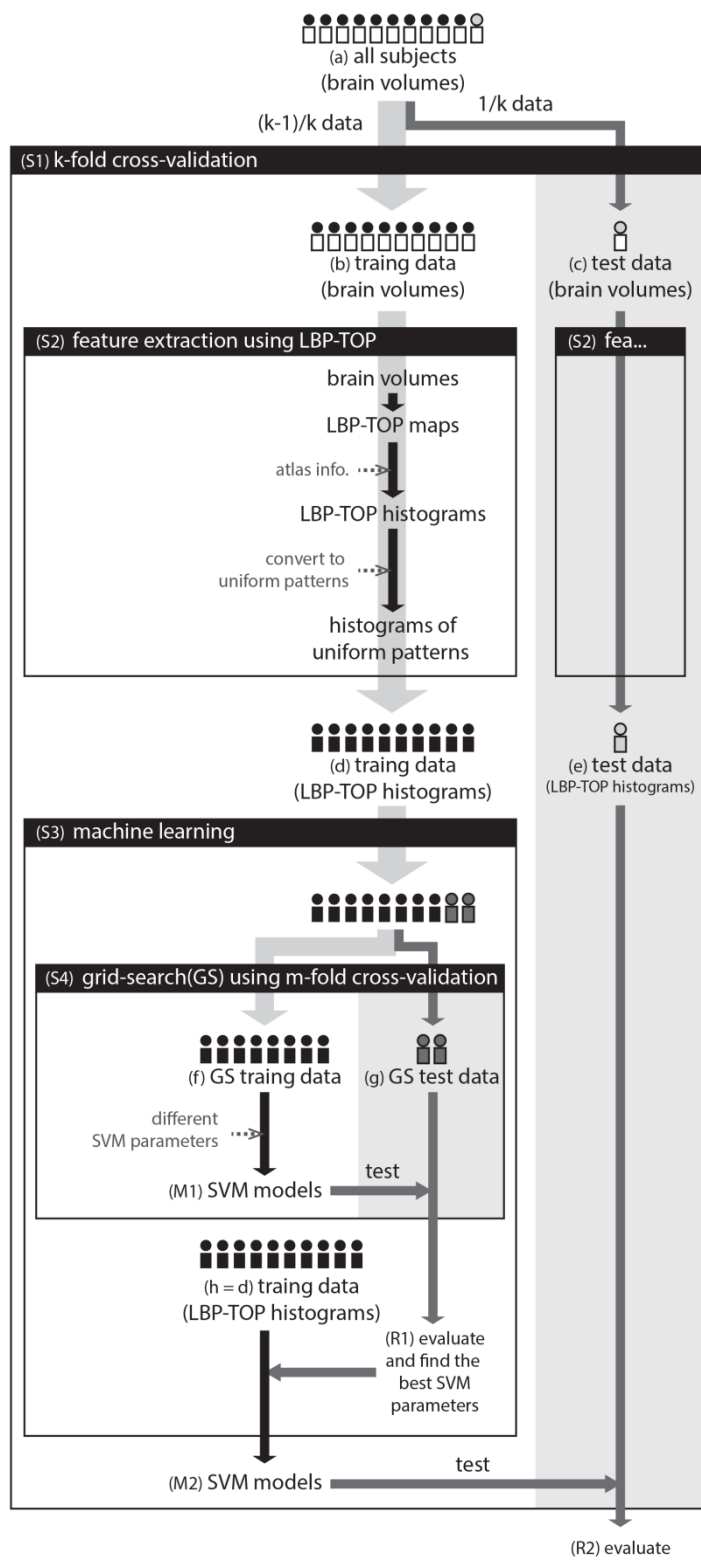


Figure 2-7 the flow diagram of the building and evaluation of a classification using SVM and LBP.



### *k-fold cross-validation (S1)*

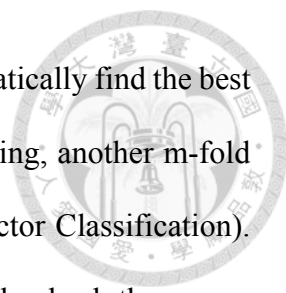
All tests in this study were evaluated by  $k$ -fold cross-validation (S1 in Figure 2-7). We randomly partitioned all subjects ((a) in Figure 2-7) into  $k$  subgroups. For each step of cross-validation, one subgroup was used as a test data set ((c) in Figure 2-7), and the remaining ( $k-1$ ) subgroups were pooled as a training data set ((b) in Figure 2-7). Then the training data set was used to train a SVM mode ((M2) in Figure 2-7). We applied the model on test data set to predict the labels or values of each subject in test data set and got the test results of each fold. After  $k$ -fold cross-validations, the test results of all  $k$  subgroups were combined to build the accuracy of the estimation of each tested model.

### *Feature extraction using LBP-TOP (S2)*

For each subject in training data set or test data set, we applied same feature extraction method to extract LBP-TOP histograms from subjects. First, the brain volumes (T1WI or parameters of DTI) were converted to LBP-TOP maps. Second, after introducing some atlas information, the LBP-TOP histograms of specific brain regions can be extracted. Third, for removing the redudent patterns, the resulting histograms were be converted to uniform patterns. The uniform patterns were the input of machine learning process.

### *Machine learning and grid-search(S3, S4)*

While using SVM, there are two main parameters should be decided ( $C$  and  $\gamma$  in RBF



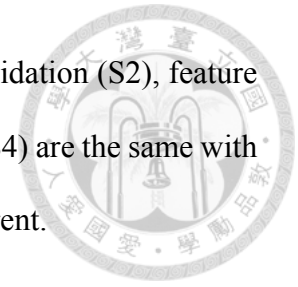
kernel). For linear SVM, only C should be decided. In order to automatically find the best parameters using in machine learning process and avoid the overfitting, another m-fold cross-validation were constructed (A Practical Guide to Support Vector Classification). As recommend in [104], we used grid-search method to automatically check the proper parameters. As showing in (S4 in Figure 2-7), different SVM parameter sets was tested and evaluated using the m-fold cross-validation. After finding the best parameters, the original training data set (h and d in Figure 2-7) were used to train the resulting classification models using the best parameter set.

Most studies in this paper used 10-fold cross-validation ( $k=10$ ) to evaluate each approaches and 5-fold cross-validation ( $m=5$ ) to automatically find the best parameters. The parameter sets using in all case are C equals 0.5, 1, 2, 4, 8, 16, 32, 64, 128, or 256 while constructing linear SVM models.

## 2.6.2 *Feature Selection and Building Regression Models Using SVR and LBP*

The flow diagram in Figure 2-8 shows the steps of feature selection in classification or regression models. While building classification models, we directly used the linear SVM implemented by LibLinear. As for constructing regression models, we used linear SVR as the feature selection tool. After selecting the top N features, SVR with RBF kernel was constructed using the same

process described in 2.6.1. In Figure 2-8, the k-fold cross-validation (S2), feature extraction (S2), and machine learning (S3), and grid-search (S4) are the same with Figure 2-7. Only the inserted feature selection steps are different.



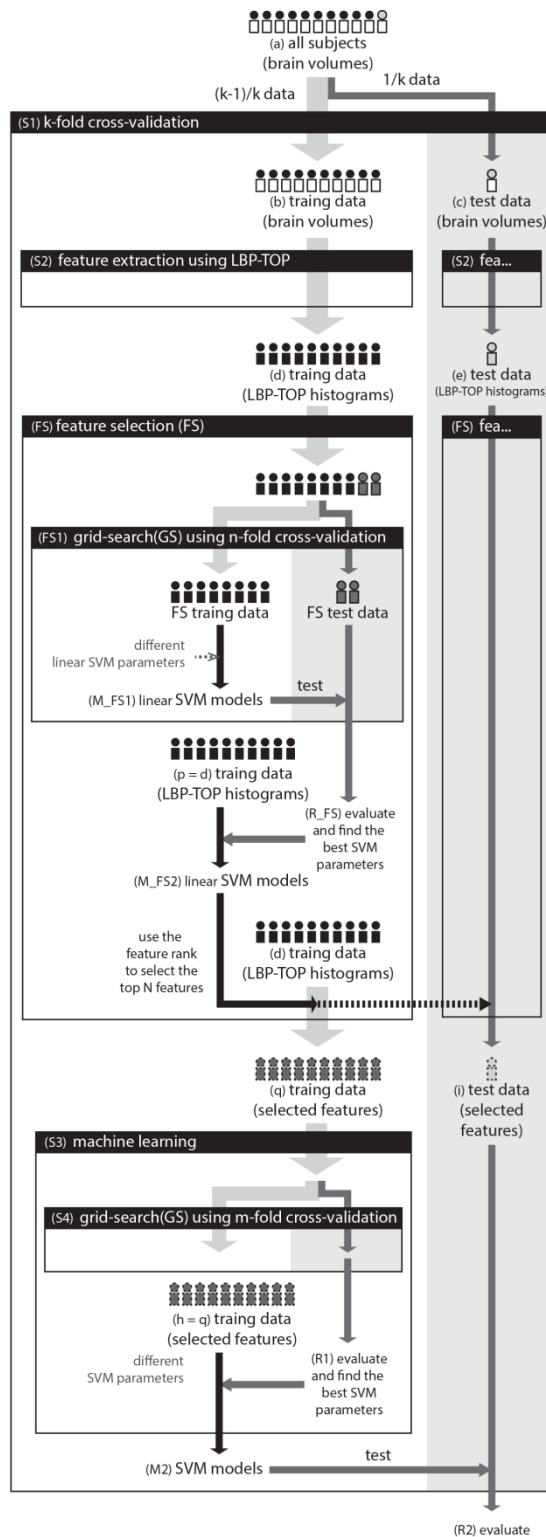


Figure 2-8 the flow diagram of feature selection in classification or regression models.

### Feature selection (FS)



LBP-TOP histograms are usually an over-complete dataset. For selecting the most important features to build simpler and more robust models, we use linear-SVM (or linear SVR) as our feature selection tool. Therefore, we should train linear SVM models using training data (d in Figure 2-8) to get the weights of each feature. Another n-fold cross-validation would be built to find the best parameters for the linear SVM model of the training data (R\_FS in in Figure 2-8). Then we used those parameters to train other linear-SVM models using the training data (p in Figure 2-8). The ranked features in the resulting linear-SVM models could be a good reference for the most important features.

Based on the ranks of features, the most N important features would be extracted from both the training data and test data. As showing in Figure 2-8, Only those N features would be the input of the machine learning process (S3 and S4 in Figure 2-8).





## Chapter 3

### Properties of LBP-TOP

In this chapter, we use the framework describing in 2.6 and basic sex classification to analyze the basic properties of LBP-TOP. There are two parts in this chapter. In first part, we analyze the algorithms and parameters of LBP-TOP. The basic parameters of LBP were tested in 3.2. Then the efficiency of uniform patterns was tested in 3.3. The alternative LBP coding methods were tested in 3.4. In second part, the properties of input brain MR images were tested. We analyzed the effects of signal-to-noise ratio (SNR) of brain MR images in 3.5. How the image resolution affects the performance of LBP-TOP was tested in 3.6. All test models were learned by linear SVM and evaluated by same 10-fold cross-validation.



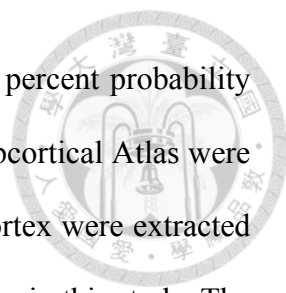
## 3.1 Materials and Methods

### 3.1.1 *Participants*

The test datasets were downloaded from a public access online database, the 1000 Functional Connectomes Project from [https://www.nitrc.org/projects/fcon\\_1000](https://www.nitrc.org/projects/fcon_1000). The 1000 Functional Connectomes Project is an unrestricted public functional MRI datasets independently collected and pooled from 33 sites. For quick tests and controlling the quality of the each dataset, only the datasets from site Cambridge\_Buckner , contributed by Dr. Randy L. Buckner, Harvard University, were used. The anatomical scans, detailed phenotypic information, and imaging parameters are all available on the website. In this test, we only used the anatomical scans acquired by magnetization-prepared rapid gradient echo (MPRAGE) sequence with voxel size =  $1.0 \times 1.0 \times 1.0$  mm. All of the available subjects were included in this study. A total 198 subjects were included, consisting of 123 females and 75 males ranging in age from 18-30years old (mean  $21.03 \pm 2.31$ ).

### 3.1.2 *Parcellations*

To obtain various scales and regions of parcellation information, three different probability atlases included with FSL (<http://www.fmrib.ox.ac.uk/fsl/>) were used in this study. These atlases include the MNI structural atlas and the Harvard-Oxford Cortical and Subcortical Atlas. The details of these atlases can be assessed on the FSL website (<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases>). The original regions in these atlases were



separated into left and right brain regions and filtered through fifty percent probability to build overlapped masks. Large regions of the Harvard-Oxford Subcortical Atlas were extracted. The left and right Cerebral White Matter and Cerebral Cortex were extracted as the white matter (WM) and gray matter (GM) regions, respectively, in this study. The lateral ventricle and brainstem regions were excluded. As a result, 2 regions in WM, 2 regions in GM, 18 regions in the MNI structural atlas (MNI), 96 regions in Harvard-Oxford Cortical Atlas (CORT), and 14 regions in Harvard-Oxford Subcortical Atlas (SUB-CORT) were used in this work.

Two sets of parcellation data were tested in this work. First, we tested the performance of using each atlas (MNI, CORT, and SUB-CORT) in the separated parcellation sets. Then, we combined MNI, CORT, SUB-CORT, GM and WM atlases to construct overlapping merged parcellation sets with different scales.

### 3.1.3 *Evaluation*

All tests in this chapter were evaluated by 10-fold cross-validation (as showing in Figure 2-7). We randomly partitioned the 198 subjects into 10 subgroups. For each step of cross-validation, one subgroup was used as a test data set, and the remaining nine subgroups were pooled as a training data set. After 10 cross-validations, the test results of all 10 subgroups were combined to build the accuracy of the estimation of each model. To facilitate comparison of the results, the same 10-fold cross-validation set was used in all evaluations.




## 3.2 Parameters of LBP

Table 3-1 shows the 10-fold cross-validation results for different radii (1 mm, 2 mm, and 3 mm) for the LBP-TOP, various parcellations, linear registrations, and non-linear registrations, respectively. Comparing the results of different radii used in LBP-TOP, we can find in the case of sex classifications, all three tested radii can provide good results. As expected, brain data with non-linear registration methods showed the highest accuracy in all cases. Particularly, using linearly registered brain data only slightly reduce accuracy. As for linear registration methods, using higher degree of freedom can provide better results.

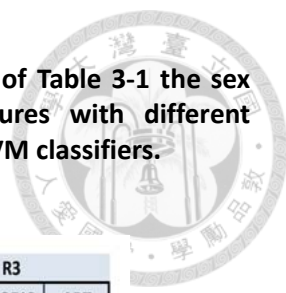
Based on the results in Table 3-1, each specific parcellation can provide information to discriminate sex, especially the brain cortical regions (CORT). However, combining information from all brain regions usually improved the resulting performance ([a], [b], and [c]). The [c] in Table 3-1 showed combining different scales of parcellations can provide stable performance across different settings.

**Table 3-1 the sex classification accuracy of models based on uniform LBP-TOP features with different registration methods, parcellation, and radius of LBP-TOP, using linear-SVM classifiers.**



parcellation		R1			R2			R3		
		DOF9	DOF12	ART	DOF9	DOF12	ART	DOF9	DOF12	ART
	GM	0.85	0.83	0.91	0.85	0.86	0.89	0.81	0.81	0.87
	WM	0.82	0.79	0.80	0.80	0.79	0.83	0.82	0.78	0.81
	CORT	0.92	0.90	0.93	0.91	0.91	0.92	0.90	0.91	0.94
	SUB_CORT	0.84	0.84	0.86	0.84	0.86	0.88	0.83	0.84	0.86
	CERE	0.85	0.87	0.89	0.88	0.88	0.91	0.87	0.87	0.92
[a]	MNI	0.90	0.89	0.96	0.91	0.89	0.95	0.91	0.90	0.93
	CORT + SUB_CORT + CERE	0.90	0.92	0.93	0.92	0.93	0.93	0.91	0.92	0.94
	[a] + GM + WM	0.90	0.91	0.93	0.91	0.93	0.93	0.91	0.92	0.94
	[a] + MNI	0.93	0.91	0.93	0.92	0.92	0.93	0.91	0.92	0.94

Abbreviations are as follows: **R1**, **R2**, and **R3**, LBP-TOP radius in mm; **DOF9**, and **DOF12**, linear registration with 9, and 12 degree of freedom, respectively; **ART**, non-linear registration performed by Automated Registration Tool; **GM**, left and right gray matter regions; **WM**, left and right white matter regions; **CORT**, the Harvard-Oxford cortical atlas; **SUB\_CORT**, the Harvard-Oxford subcortical atlas; **CERE**, the probabilistic cerebellar atlas; **MNI**, the MNI structural atlas. All atlases were provided by FSLView version 3.0[105]. The sensitivity, specificity, and areas under the ROC curve (AUC) of this table can be found in Table 3-2.



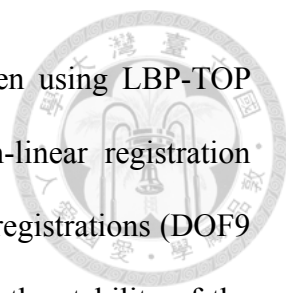
**Table 3-2 Sensitivity, specificity, and areas under the ROC curve (AUC) of Table 3-1 the sex classification accuracy of models based on uniform LBP-TOP features with different registration methods, parcellation, and radius of LBP-TOP, using linear-SVM classifiers.**

Sensitivity		R1			R2			R3		
		DOF9	DOF12	ART	DOF9	DOF12	ART	DOF9	DOF12	ART
	parcellation									
	GM	0.81	0.80	0.91	0.83	0.84	0.83	0.77	0.79	0.83
	WM	0.75	0.75	0.72	0.73	0.69	0.77	0.77	0.68	0.73
	CORT	0.87	0.85	0.88	0.85	0.85	0.87	0.84	0.84	0.89
	SUB_CORT	0.75	0.72	0.79	0.75	0.76	0.81	0.71	0.77	0.77
	CERE	0.76	0.80	0.83	0.79	0.85	0.87	0.77	0.81	0.87
	MNI	0.87	0.84	0.93	0.87	0.89	0.92	0.88	0.89	0.92
[a]	CORT + SUB_CORT + CERE	0.87	0.88	0.88	0.89	0.89	0.89	0.85	0.87	0.91
[b]	[a] + GM + WM	0.87	0.88	0.89	0.88	0.89	0.89	0.85	0.87	0.91
[c]	[b] + MNI	0.89	0.87	0.89	0.89	0.88	0.89	0.87	0.87	0.91

Specificity		R1			R2			R3		
		DOF9	DOF12	ART	DOF9	DOF12	ART	DOF9	DOF12	ART
	parcellation									
	GM	0.88	0.85	0.92	0.86	0.87	0.93	0.84	0.82	0.89
	WM	0.87	0.82	0.85	0.84	0.85	0.87	0.85	0.84	0.85
	CORT	0.95	0.93	0.96	0.94	0.94	0.96	0.94	0.95	0.97
	SUB_CORT	0.89	0.91	0.91	0.89	0.92	0.92	0.90	0.89	0.92
	CERE	0.91	0.91	0.93	0.94	0.90	0.94	0.93	0.91	0.95
	MNI	0.93	0.93	0.98	0.94	0.89	0.98	0.93	0.91	0.94
[a]	CORT + SUB_CORT + CERE	0.93	0.94	0.97	0.94	0.96	0.95	0.94	0.96	0.97
[b]	[a] + GM + WM	0.93	0.93	0.96	0.93	0.95	0.95	0.94	0.96	0.97
[c]	[b] + MNI	0.95	0.94	0.96	0.93	0.94	0.96	0.94	0.95	0.96

AUC		R1			R2			R3		
		DOF9	DOF12	ART	DOF9	DOF12	ART	DOF9	DOF12	ART
	parcellation									
	GM	0.85	0.82	0.91	0.84	0.85	0.88	0.81	0.80	0.86
	WM	0.81	0.78	0.78	0.79	0.77	0.82	0.81	0.76	0.79
	CORT	0.91	0.89	0.92	0.90	0.90	0.91	0.89	0.90	0.93
	SUB_CORT	0.82	0.82	0.85	0.82	0.84	0.87	0.80	0.83	0.85
	CERE	0.84	0.86	0.88	0.86	0.88	0.90	0.85	0.86	0.91
	MNI	0.90	0.88	0.95	0.90	0.89	0.95	0.90	0.90	0.93
[a]	CORT + SUB_CORT + CERE	0.90	0.91	0.92	0.92	0.93	0.92	0.90	0.91	0.94
[b]	[a] + GM + WM	0.90	0.90	0.93	0.91	0.92	0.92	0.90	0.91	0.94
[c]	[b] + MNI	0.92	0.90	0.93	0.91	0.91	0.93	0.90	0.91	0.93

Abbreviations are as follows: **R1**, **R2**, and **R3**, LBP-TOP radius in mm; **DOF9**, and **DOF12**, linear registration with 9, and 12 degree of freedom, respectively; **ART**, non-linear registration performed by Automated Registration Tool; **GM**, left and right gray matter regions; **WM**, left and right white matter regions; **CORT**, the Harvard-Oxford cortical atlas; **SUB\_CORT**, the Harvard-Oxford subcortical atlas; **CERE**, the probabilistic cerebellar atlas; **MNI**, the MNI structural atlas. All atlases were provided by FSLView version 3.0[105]. The AUC was calculated using [106] and the positive subject is male subjects.



Another finding is the robustness of the registration methods when using LBP-TOP features to classify sex. Although the models based on the non-linear registration method are most accurate in each cases, the models based on linear registrations (DOF9 and DOF12) also performed well (Table 3-1). This finding shows the stability of the LBP-TOP to registration methods. Therefore, we also test the non-normalized brain MR images approaches in Chapter 4.

### **3.3 The Efficiency of Uniform Patterns**

Table 3-3 shows the effect of LBP-TOP features with and without converting to uniform patterns. In each case of different radii, registration methods, and parcellations, there are performance difference between approaches using uniform patterns or not. However, the resulting feature size of using non-uniform patterns is more than four times bigger than introducing uniform patterns. The bigger feature size something hurts the resulting accuracy in some complex machine learning approaches. Moreover, much more features can limit the size of training subjects in same computational resources and then take much time to learn models from same subject size. This test results show the efficiency of using uniform patterns. Therefore, we always used uniform patterns in this study.

**Table 3-3 the sex classification accuracy of models based on LBP-TOP features with and without converting to uniform patterns.**

parcellation	uniform	features	R1			R2			R3		
			DOF9	DOF12	ART	DOF9	DOF12	ART	DOF9	DOF12	ART
CORT	non-uniform	73728	0.90	0.86	0.94	0.89	0.89	0.93	0.90	0.90	0.94
	uniform	16992	0.92	0.90	0.93	0.91	0.91	0.92	0.90	0.91	0.94
MNI	non-uniform	13824	0.90	0.91	0.94	0.93	0.91	0.95	0.92	0.92	0.95
	uniform	3186	0.90	0.89	0.96	0.91	0.89	0.95	0.91	0.90	0.93
[c]	non-uniform	122880	0.91	0.90	0.93	0.93	0.93	0.94	0.92	0.92	0.94
	uniform	28320	0.93	0.91	0.93	0.92	0.92	0.93	0.91	0.92	0.94

Abbreviations are as follows: **R1**, **R2**, and **R3**, LBP-TOP radius in mm; **DOF9**, and **DOF12**, linear registration with 9, and 12 degree of freedom, respectively; **ART**, non-linear registration performed by Automated Registration Tool; **CORT**, the Harvard-Oxford cortical atlas; **MNI**, the MNI structural atlas; **[c]**, the same parcellations used in Table 3-1 [c]. All atlases were provided by FSLView version 3.0[105].

### 3.4 The Order of LBP Coding

In this section, we showed the order of LBP coding is arbitrary and do not affect the results of SVM models. We modified the coding series of LBP (weights in the step II-2 in Figure 2-2) histogram. As showing in Figure 3-1, the resulting LBP maps of different coding orders are very different. However, the linear SVM would always build same classification model using same features with same ranks. Therefore, as the results in Table 3-4, the resulting accuracies of every orders are the same.



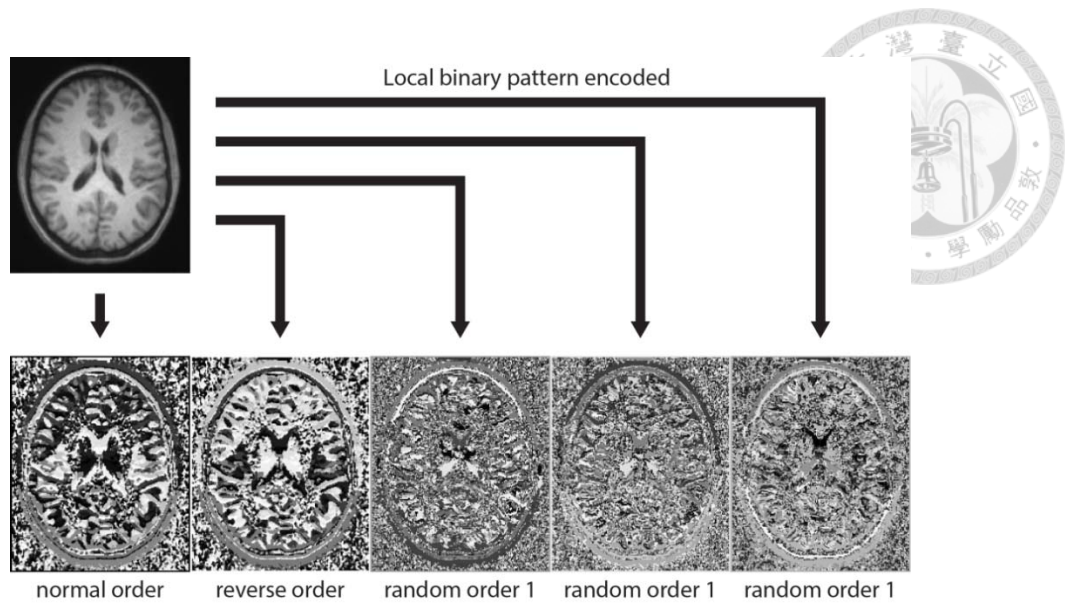


Figure 3-1 LBP map using different coding series.

Table 3-4 the sex classification accuracy of models based on LBP-TOP features with different encode methods, 1mm resolution, 2mm radius, non-linear registration, and uniform patterns.

parcellation	encode method	ART
CORT	normal order	0.92
	reverse order	0.92
	random order 1	0.92
	random order 2	0.92
	random order 3	0.92
MNI	normal order	0.95
	reverse order	0.95
	random order 1	0.95
	random order 2	0.95
	random order 3	0.95
[c]	normal order	0.93
	reverse order	0.93
	random order 1	0.93
	random order 2	0.93
	random order 3	0.93

Abbreviations are as follows: **ART**, non-linear registration performed by Automated Registration Tool; **CORT**, the Harvard-Oxford cortical atlas; **MNI**, the MNI structural atlas; **[c]**, the same parcellations used in Table 3-1 [c]. All atlases were provided by FSLView version 3.0[105].

### 3.5 The Effects of Brain MR Image SNR



To simply test the effects of image SNR to the performance of LBP-TOP approaches, we add Gaussian noise to the original brain MR images. The signals of the images were estimated by the average of 80% white matter region based on the probabilistic atlas provided by FSLView version 3.0 [105]. In each level of SNR, we add Gaussian noise with zero mean and the standard deviation based on the signal and SNR ( $\text{SNR} = \text{signal} / \text{standard deviation}$ ). Figure 3-2 shows the examples of the brain MR images and corresponding LBP maps in every tested SNR levels.

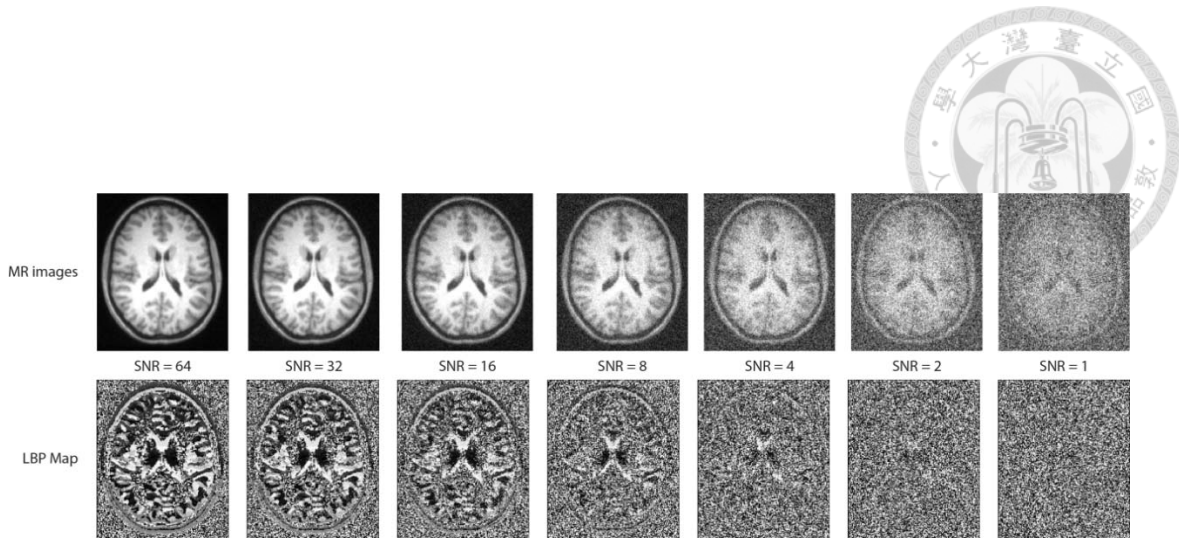
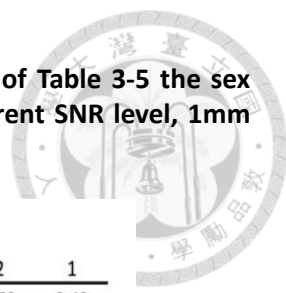


Figure 3-2 brain MR images and the corresponding LBP map in different SNR levels.

Table 3-5 the sex classification accuracy of models based on LBP-TOP features with different SNR level, 1mm resolution, 2mm radius, non-linear registration, and uniform patterns.

parcellation		SNR							
		64	32	16	8	4	2	1	
[a]	GM	0.87	0.85	0.82	0.78	0.70	0.67	0.58	
	WM	0.80	0.74	0.72	0.63	0.65	0.57	0.54	
	CORT	0.93	0.92	0.92	0.92	0.89	0.81	0.74	
	SUB_CORT	0.88	0.89	0.85	0.77	0.74	0.71	0.57	
	CERE	0.91	0.85	0.83	0.82	0.80	0.76	0.60	
	MNI	0.93	0.94	0.93	0.91	0.84	0.81	0.72	
	CORT + SUB_CORT + CERE	0.94	0.95	0.95	0.93	0.88	0.82	0.72	
	[a] + GM + WM	0.93	0.94	0.94	0.92	0.89	0.82	0.72	
	[b]	[a] + MNI	0.94	0.94	0.94	0.94	0.89	0.83	0.73
	[c]	[b] + MNI	0.94	0.94	0.94	0.94	0.89	0.83	0.73

Abbreviations are as follows: **GM**, left and right gray matter regions; **WM**, left and right white matter regions; **CORT**, the Harvard-Oxford cortical atlas; **SUB\_CORT**, the Harvard-Oxford subcortical atlas; **CERE**, the probabilistic cerebellar atlas; **MNI**, the MNI structural atlas. All atlases were provided by FSLView version 3.0[105]. The sensitivity, specificity, and areas under the ROC curve (AUC) of this table can be found in Table 3-6.



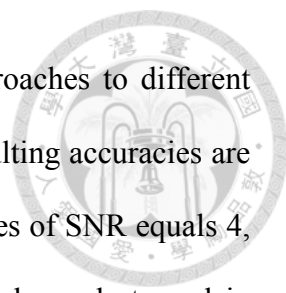
**Table 3-6 Sensitivity, specificity, and areas under the ROC curve (AUC) of Table 3-5 the sex classification accuracy of models based on LBP-TOP features with different SNR level, 1mm resolution, 2mm radius, non-linear registration, and uniform patterns.**

		Sensitivity		SNR					
		parcellation		64	32	16	8	4	2
		GM	0.85	0.80	0.75	0.72	0.57	0.53	0.43
		WM	0.71	0.63	0.65	0.44	0.49	0.40	0.32
		CORT	0.89	0.89	0.88	0.88	0.79	0.61	0.35
		SUB_CORT	0.81	0.81	0.75	0.61	0.52	0.43	0.19
		CERE	0.85	0.77	0.76	0.71	0.67	0.52	0.11
		MNI	0.89	0.91	0.91	0.85	0.73	0.64	0.47
[a]		CORT + SUB_CORT + CERE	0.91	0.91	0.92	0.87	0.77	0.60	0.27
[b]		[a] + GM + WM	0.89	0.91	0.91	0.87	0.79	0.59	0.28
[c]		[b] + MNI	0.91	0.91	0.91	0.91	0.80	0.63	0.31

		Specificity		SNR					
		parcellation		64	32	16	8	4	2
		GM	0.89	0.88	0.86	0.82	0.77	0.75	0.67
		WM	0.85	0.80	0.76	0.74	0.75	0.67	0.67
		CORT	0.96	0.94	0.94	0.95	0.95	0.93	0.98
		SUB_CORT	0.92	0.93	0.91	0.87	0.88	0.88	0.80
		CERE	0.95	0.89	0.88	0.89	0.89	0.91	0.89
		MNI	0.96	0.96	0.94	0.94	0.91	0.91	0.87
[a]		CORT + SUB_CORT + CERE	0.96	0.98	0.97	0.97	0.95	0.96	0.99
[b]		[a] + GM + WM	0.95	0.96	0.97	0.96	0.95	0.96	0.99
[c]		[b] + MNI	0.96	0.96	0.96	0.97	0.95	0.95	0.98

		AUC		SNR					
		parcellation		64	32	16	8	4	2
		GM	0.87	0.84	0.80	0.77	0.67	0.64	0.55
		WM	0.78	0.72	0.71	0.59	0.62	0.53	0.49
		CORT	0.93	0.92	0.91	0.92	0.87	0.77	0.66
		SUB_CORT	0.87	0.87	0.83	0.74	0.70	0.65	0.49
		CERE	0.90	0.83	0.82	0.80	0.78	0.72	0.50
		MNI	0.93	0.93	0.92	0.90	0.82	0.78	0.67
[a]		CORT + SUB_CORT + CERE	0.93	0.94	0.94	0.92	0.86	0.78	0.63
[b]		[a] + GM + WM	0.92	0.93	0.94	0.91	0.87	0.77	0.64
[c]		[b] + MNI	0.93	0.93	0.93	0.94	0.88	0.79	0.65

Abbreviations are as follows: **GM**, left and right gray matter regions; **WM**, left and right white matter regions; **CORT**, the Harvard-Oxford cortical atlas; **SUB\_CORT**, the Harvard-Oxford subcortical atlas; **CERE**, the probabilistic cerebellar atlas; **MNI**, the MNI structural atlas. All atlases were provided by FSLView version 3.0[105]. The AUC was calculated using [106] and the positive subject is male subjects.

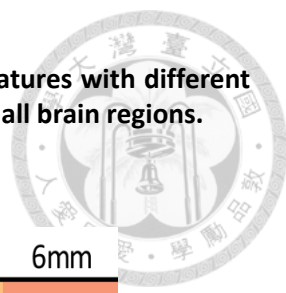


Results in Table 3-5 demonstrate the robustness of LBP-TOP approaches to different SNR levels. While using images with higher SNR (8 to 64), the resulting accuracies are the same based on informations from every brain regions. In the cases of SNR equals 4, although we can find the edges and structures of LBP maps have been destroyed in Figure 3-2, the LBP-TOP approaches perform well. The LBP histograms in different brain regions still show sufficient information to machine learning algorithms to classify sex. In the cases of SNR equals to 2, there are no edges or structures can be found in LBP maps (Figure 3-2). However, the LBP histograms can provide some useful information to classify the sex. Only while the SNR equals to 1, the machine learning algorithms learned almost nothing form the LBP histograms.

### **3.6 The Effects of Brain MR Image Resolution**

**Table 3-7** shows the accuracy of models based on various brain image resolutions. Although the resolution does not affect the results too much, data with higher resolution generally provide more information for the discrimination of subject sex. In our experience, the effects of resolution are case by case. Higher-resolution images are usually better than low ones. However, there is no robust rule for sufficient resolution for every approach.

**Table 3-7 the sex classification accuracy of models based on LBP-TOP features with different resolution, radius, registration method, uniform patterns, and combined all brain regions.**



registration	radius	1mm	2mm	3mm	4mm	6mm
DOF12	R1	0.91	0.92	0.90	0.88	0.84
	R2	0.92	0.93	0.92	0.92	0.92
	R3	0.92	0.94	0.92	0.92	0.92
ART	R1	0.93	0.95	0.93	0.90	0.88
	R2	0.93	0.93	0.92	0.91	0.93
	R3	0.94	0.93	0.94	0.94	0.93

Abbreviations are as follows: **R1**, **R2**, and **R3**, LBP-TOP radius in mm; **DOF12**, linear registration with 12 degree of freedom, respectively; **ART**, non-linear registration performed by Automated Registration Tool.

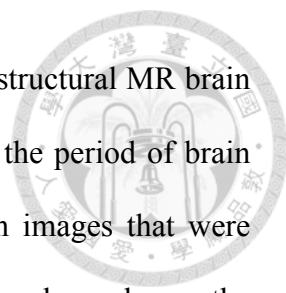


## **Chapter 4**

# **Age Estimation Using Unnormalized MR Brain Images**

### **4.1 Introduction**

Aging-related brain morphological change is an important issue in neuroscience. Many studies have shown cerebral changes during brain maturation and aging [37-44]. To illustrate cognitive development and decline, or to explore aging-related mental disorders based on changes in brain morphology, an age biomarker defining the normal patterns of structural brain changes over the lifespan is critical.

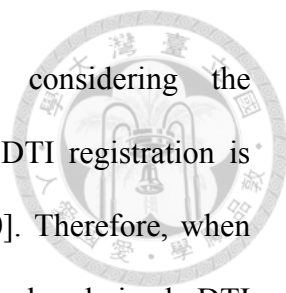


Recently, a variety of methods were proposed to estimate age using structural MR brain images [107-112]. However, most of these studies only considered the period of brain maturation or aging. Moreover, the studies all used acquired brain images that were transformed to a standard template and compared voxel-by-voxel, such as the voxel-based morphometry (VBM) analysis [50, 52]. However, several unsolved problems exist in this approach (see 0 for more details).

In this study, to avoid all the mentioned risks, we used a simple feature extraction method based on local binary patterns on three orthogonal planes (LBP-TOP) to extract effective information from MR brain images without registration [73, 80, 83]. In the first part of this chapter, we used LBP-TOP to extract brain morphological change information from unnormalized brain images, which provide pure and unmodified raw information. Furthermore, most studies only use gray matter (GM) segmentation as source data. The process of segmenting brain images into GM, white matter (WM), and cerebrospinal fluid (CSF) could be another step that restricts the analysis pipeline and introduces noise to the image data [113-115]. In addition to using GM segmentations, we also used whole-brain images as an information source.

Diffusion tensor imaging (DTI) is an MRI technique that measures the restricted diffusion of water molecules in tissue and offers directional information of neural tracts in the brain [8]. DTI has proven its value in WM mapping and anatomical connectivity in brain research as well as clinical applications [9, 10]. Moreover, many studies show that DTI can reveal microstructural changes in brain tissue during brain maturation and aging [116, 117]. Recently, Benson Mwangi et al. have revealed the possibility of

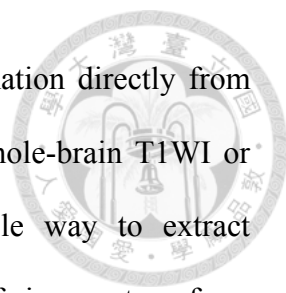




predicting age using DTI measurements [111]. However, considering the multi-dimensional nature of the data and the tensor orientations, DTI registration is much more challenging compared to using scalar images [118-120]. Therefore, when analyzing DTI data, most studies considered only an eigenvalue-derived DTI measurement, such as fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD), and so on. Few of these studies used the information in fiber spatial orientation (eigenvectors) [117]. Inspired by the recent research of Wedeen et al., which showed that the cerebral fiber pathways form a regular grid based on the three principal axes of development [121], we used the orientation of fiber to estimate age in this study. The capability of LBP-TOP to extract information from unnormalized brain images makes it an excellent candidate for analyzing unnormalized DTI data. In the second part of this chapter, we tried to use the information derived from DTI eigenvalues (FA, MD, or AD) as well as the direction of the eigenvectors to predict brain age.

In previous chapter, we used LBP-TOP to compute features from registered brain T1-weighted images (T1WI) to sort ADHD patients from normal subjects [122]. In this chapter, the method was used to estimate age with unnormalized brain T1WI or DTI measurements. To evaluate the performance of LBP-TOP in brain age estimation, we use linear support vector regression (linear SVR) as a simple feature selection method to find the most useful features for predicting brain age. Then, we used those features to train a non-linear SVR with a radial basis function (RBF) kernel for more accurate age estimation.

This study provides three contributions. First, we enable unnormalized MR brain image



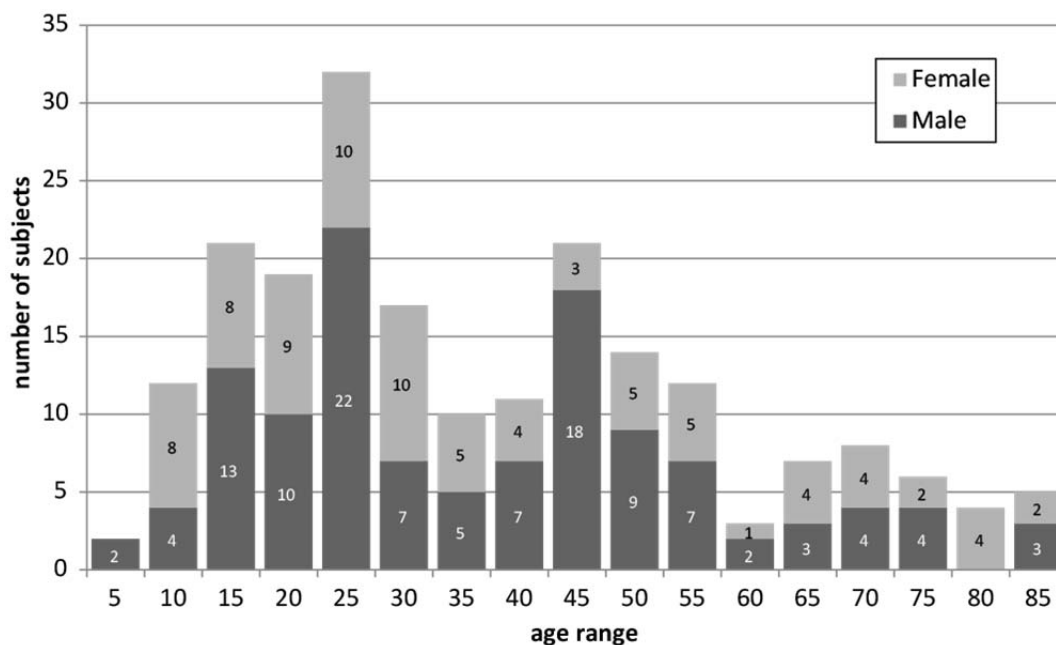
analysis by introducing LBP-TOP to extract morphological information directly from acquired images. Second, we built age estimation models from whole-brain T1WI or GM segments using LBP-TOP. Third, we demonstrate a simple way to extract discriminative information from eigenvalues, or even the direction of eigenvectors, from DTI data. To the best of our knowledge, no previous study has used unnormalized MR brain images to predict brain age. More importantly, this may be the first report that the information encoded in fiber orientation could be used to estimate the brain age.

## 4.2 Materials and Methods

### 4.2.1 *Subjects and Data Acquisition*

The employed datasets were downloaded from a public access online database, the Nathan Kline Institute (NKI) / Rockland Sample from the International Neuroimaging Data-sharing Initiative (INDI). The NKI/Rockland Sample is a phenotypically rich neuroimaging sample for discovery science. The anatomical scans, DTI scans, detailed phenotypic information, and imaging parameters are all available on the website ([http://fcon\\_1000.projects.nitrc.org/indi/pro/nki.html](http://fcon_1000.projects.nitrc.org/indi/pro/nki.html)). In this study, we used the anatomical scans and 64-direction DTI scans from the database. Both scans were acquired using a 3T Siemens Trio scanner. The anatomical scans were acquired by magnetization-prepared rapid gradient echo (MPRAGE) sequence with TR/TE =

2500/3.5 ms and voxel size = 1.0×1.0×1.0 mm. The 64-direction DTI scans were acquired by EPI sequence with TR/TE = 10000/91 ms, voxel size = 2.0×2.0×2.0 mm, and b-value = 1000 s/mm<sup>2</sup>. All of the available subjects were included in this study, with the exception of three subjects without MPRAGE scans (subject id: 1933343, 2136756, and 2479362). Therefore, a total 204 subjects were included, consisting of 84 females and 120 males ranging in age from 4-85 years old (mean 35.22 ± 25.07). Figure 4-1 shows the age distribution of the subjects.



**Figure 4-1 Age distribution of the 204 subjects that participated in the study.**

The age of each subject can be found on the NKI/Rockland Sample website ([http://fcon\\_1000.projects.nitrc.org/indi/pro/nki.html](http://fcon_1000.projects.nitrc.org/indi/pro/nki.html)).



#### 4.2.2 *Data preprocessing*

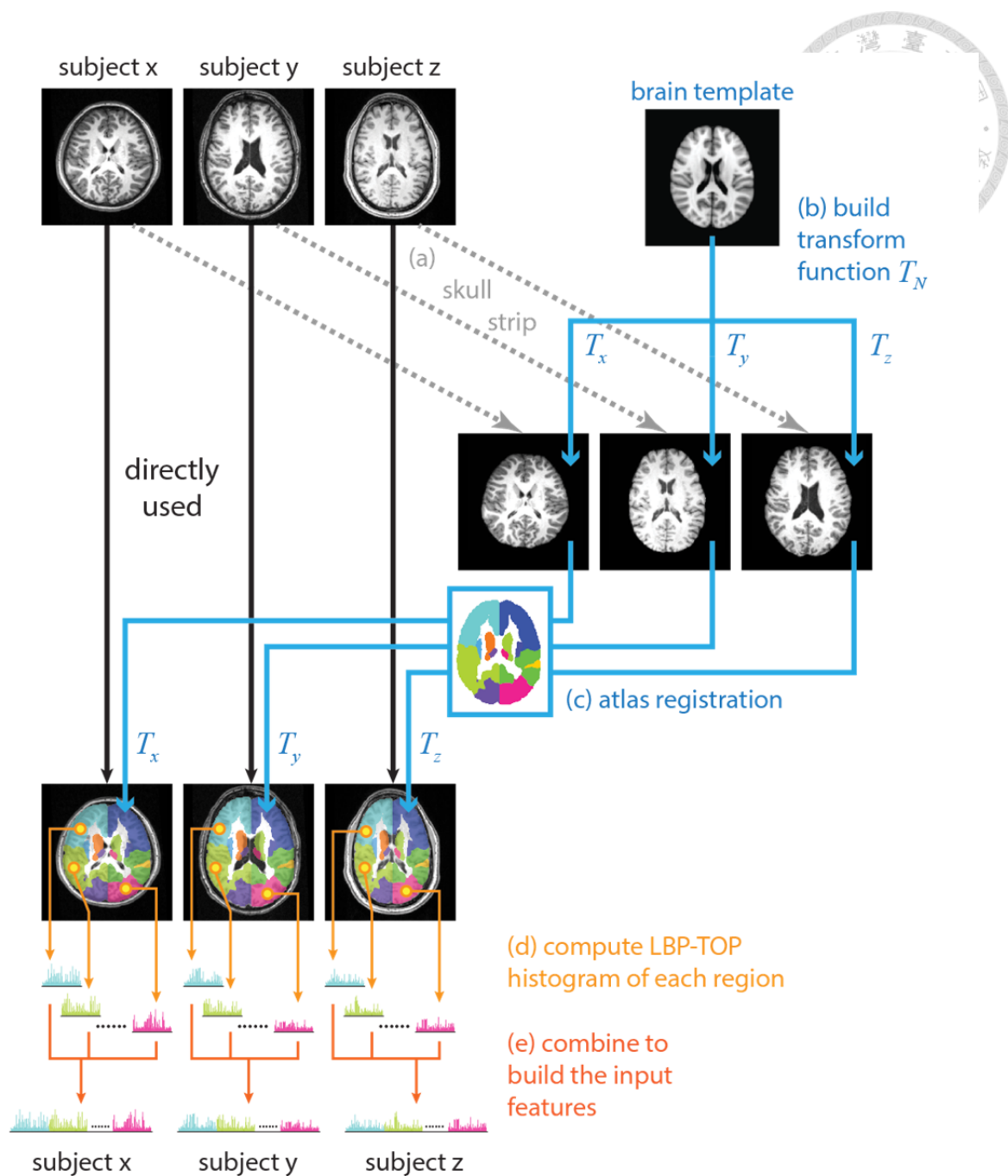
All of the images were preprocessed using FSL version 4.1.7 [123]. For T1WI, brain segmentation was performed by FSL's automated segmentation toolbox (FAST). The skull-stripped brain images were segmented into gray matter (GM), white matter (WM) and CSF [123, 124] probability maps. A 0.25 threshold was applied on the GM segment results to remove noise. Only the source T1WI head images and GM segment data were used in this work.

The DTI images were preprocessed through the following pipeline: (1) eddy current correction with FSL `eddy_correct`, (2) skull removal and brain extraction using FSL BET [125], (3) diffusion tensor model fitting of each voxel using FSL `dtifit`. Next, the raw output T2 signal images (S0) were used as the reference for atlas registration. The fractional anisotropy images (FA), mean diffusivity images (MD), first eigenvalue images (L1, known as axial diffusivity, AD), or first eigenvector data (V1) were used for age estimation.

#### 4.2.3 *Extracting the LBP-TOP Histogram*

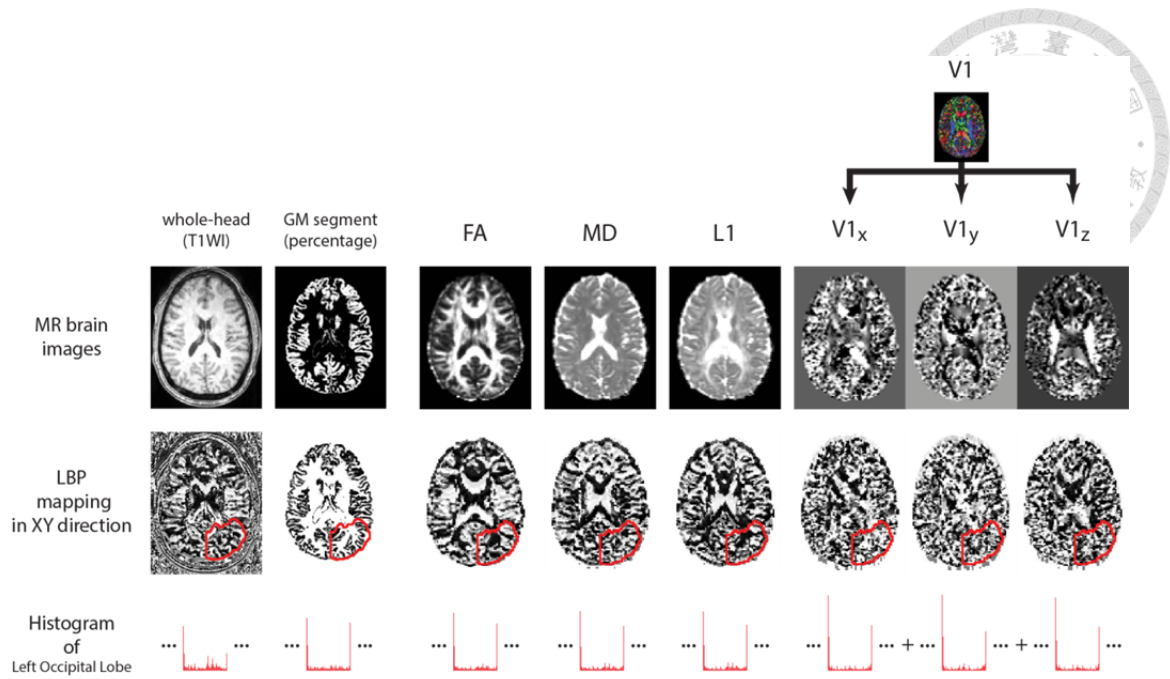
All of the acquired head images were first aligned to the same orientation by rigid transformation. Then, we extracted the LBP-TOP histograms from those head images. Figure 4-2 shows the basic steps of extracting the LBP-TOP features. The input features of the SVR were the combined histograms from the brain regions. To estimate age using T1WI, the MNI template provided in FSLView version 3.0 (file name: `MNI152_T1_1mm_brain.nii.gz`) was used as the standard template. To predict age using

the DTI data, the T2-weighted ICBM 2009c nonlinear asymmetric template was used as standard template [126, 127] for the S0 images. Next, the transformed atlas information was applied on all of the DTI measurement data. To address the first eigenvector data (V1), we compute three LBP-TOP maps of three directions to extract the information from the vector data (Figure 4-3 shows an example).



**Figure 4-2** Flow chart for extracting the LBP-TOP histogram from unnormalized MR brain images.

(a) For better atlas registration results, we first performed skull-stripping to acquire the brain image. (b) Then, using the linear or non-linear method, the transform function of atlas registration was computed to register the standard brain template to each brain image. (c) The transform function was used to transform the atlas to each subject's head image. (d) The LBP-TOP histogram of each brain region can be computed based on the atlas. (e) The combined histogram from all of the brain regions was the input feature to SVR for age estimation.



**Figure 4-3 Example LBP-TOP maps of each tested MR brain image data in this study.**

Only the LBP-TOP from the XY direction with a radius equal to two voxels is shown. The images in row 1 are the example MR brain images without registration. The images in row 2 show the XY direction LBP mapping of each image in row 1. Row 3 denotes the XY direction LBP histograms of the 3D left occipital lobe region, one of the 18 MNI regions (shown in the row 2 images as a red region) of each image. L1 denotes the first eigenvalues of the DTI data. V1 denotes the first eigenvectors of the DTI data.

#### 4.2.4 Atlas Registration

To introduce spatial information of the brain regions in this study, the following 3 probability atlases provided by FSLView version 3.0 were used: the MNI structural atlas (MNI) [128], the Harvard-Oxford cortical atlas (CORT) and the Harvard-Oxford subcortical atlas (SUB\_CORT), and the probabilistic cerebellar atlas (CERE)[129]. The

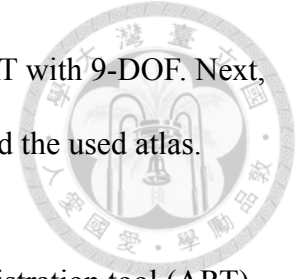
details of these atlases can be found on the FSL website (<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases>). Regions in the atlases were separated into left and right brain portions and filtered by zero percent probability to build overlapped masks. Some of the large regions from the Harvard-Oxford subcortical atlas were extracted. The left and right cerebral white matter and cerebral cortex were extracted as the gray matter (GM) and white matter (WM) regions. The lateral ventricle and brain stem regions were excluded. As a result, there were 2 regions in the WM, 2 regions in the GM, 18 regions in the MNI structural atlas (MNI), 96 regions in the Harvard-Oxford cortical atlas (CORT), 14 regions in the Harvard-Oxford subcortical atlas (SUB-CORT), and 28 regions in the probabilistic cerebellar atlas (CERE).

Three sets of atlases were tested. First, the GM and WM were used to test the information encoded in the whole-brain LBP-TOP distribution. Second, CORT, SUB\_CORT, and CERE were used to evaluate the age estimation performance of different brain segments. Third, two atlases with different scale and details were used to extract features from all of the brain regions. MNI provides large and rough spatial information from the whole-brain. Lastly, we combined all of the above mentioned atlases to build a mixed atlas with multiple scale regions. That is, GM+WM+CORT+SUB\_CORT+CERE+MNI.

Both linear and non-linear atlas registration methods were evaluated in this study. Linear registrations were performed using the linear multimodality registration method developed by FSL FLIRT [123, 130]. The brain template was first transformed to each subject brain image by FLIRT with 6 degree of freedom (6-DOF, rigid-body



transformation). The results were then linearly transformed by FLIRT with 9-DOF. Next, the results were linearly transformed by FLIRT with 12-DOF to build the used atlas.

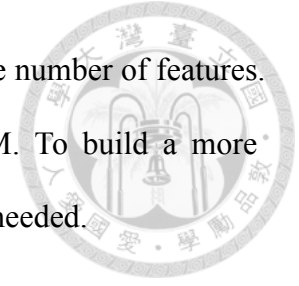


Non-linear atlas registration was performed using the automated registration tool (ART). ART was developed by Ardekani et al. [131] and can be downloaded from <http://www.nitrc.org/projects/art/>. Arno Klein et al. demonstrated that ART provides better efficiency and consistency than other non-linear registration methods [60].

#### 4.2.5 *Support Vector Regression and Feature Ranking*

Support vector regression (SVR) was used to build an age estimation model. The idea of SVR is derived from support vector machine (SVM) [132, 133]. SVM maps training data into high-dimensional feature space to find the separating hyperplane with the maximal margin. The support vectors are the data points lying closest to the hyperplane. For SVR, we need to find a function that fits as many data points as possible. Therefore, the regression line is surrounded by a tube [99, 134]. The regression line best fits those points in the tube, while points outside of the tube are the training errors. The SVR support vectors are the data points lying closest to the edge of the tube. To determine the width of the tube, we used  $\nu$ -SVR in this study. Namely, we specified an upper bound  $\nu$  on the fraction of points lying outside of the tube and automatically adjusted the width of the tube. In this study, we used LIBSVM to perform SVR with a linear kernel and a radial basis function (RBF) kernel [135].

The LBP-TOP histogram is an over-complete feature set with a huge number of features. Therefore, the histogram is not suitable for training a RBF SVM. To build a more efficient and robust regression model, a feature selection method is needed.

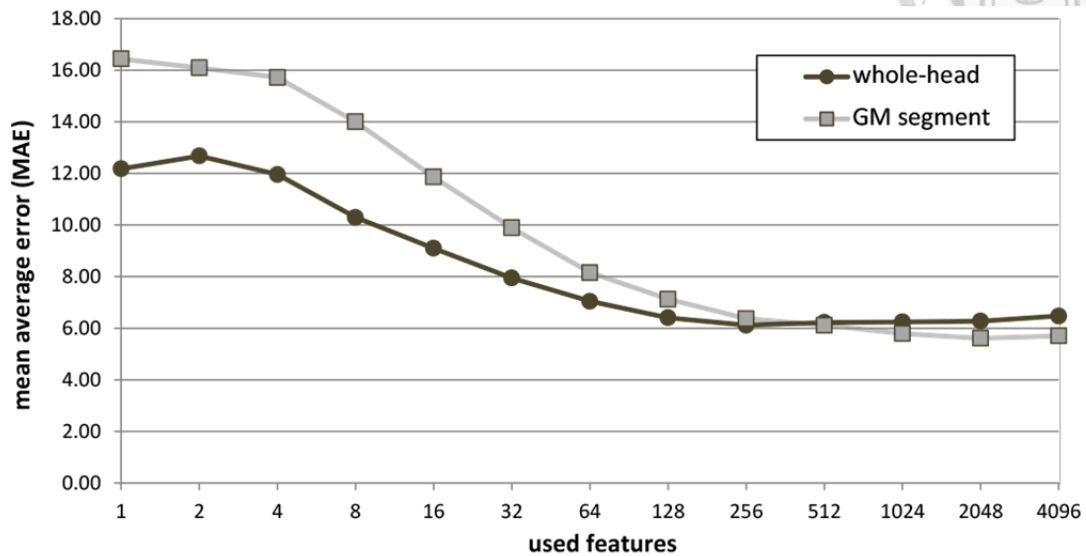


Feature ranking and selection based on a linear SVM has proven to be efficient and useful for gene selection, document classification and many other applications [102, 136, 137]. In this study, we used the same idea to select the most important bins in the LBP-TOP histogram. For any test subject, the estimated value based on the linear SVR is

$$f(x) = w^T x + b \quad (4-1)$$

where  $x$  is the feature vector,  $b$  is a constant, and  $w$  is the weight vector. Each value of  $w$  denotes the weight of each feature to compute the estimated result. The larger the absolute value of  $w_j$ , the more important the  $j$ th feature is in deciding the resulting value. After training a linear SVR model, the  $w$  in (3-1) can be used as a relative importance index. Therefore, we can build a simpler model using the top  $N$  important features.

For each test in this study, in each fold of cross-validation, we first trained a linear SVR model using the training subjects and then ranked the features. Next, we used the top  $N$  features to train a non-linear SVR model with an RBF kernel (RBF SVR) using the same training subjects. Then, we tested the RBF SVR model using the tested subjects to evaluate the test performance. Figure 4-4 shows an example feature selection result.



**Figure 4-4 Example of feature selection using linear SVR with whole-brain or GM segment data.**

The figure shows the mean absolute error (MAE) from using RBF SVR to train the top N ranked features selected by linear SVR. The selected LBP-TOP features are built by linear atlas registration, with a voxel radius of one, and combined brain regions (GM+WM+CORT+SUB\_CORT+CERE+MNI in )

#### 4.2.6 Evaluations

All of the tests in this study were evaluated by 10-fold cross-validation. The 204 subjects were randomly partitioned into 10 subgroups. For each step of cross-validation, one subgroup was used as a test data set, and the remaining nine subgroups were pooled as a training data set. After 10 cross-validations, the test results of all 10 subgroups were combined to calculate the estimation error of each test. To facilitate a comparison of the

results, the same 10-fold cross-validation set was used in all of the evaluations.



### 4.3 Results and Discussion

Figure 4-3 shows an example of LBP-TOP of the XY direction. Figure 4-4 shows an example of feature selection by linear SVR. In most of the tests we attempted in this study, we found that using the top 2048 ranked features usually provided the best age estimation model. When using more than 2048 features, the performance did not improve. Therefore, the results we show in this paper were mainly from using the top 2048 ranked features. Figure 4-4 also shows the different results from using the whole-brain or GM segment data. Although they both end up with a similar mean absolute error (MAE) after using more than 256 ranked features, the whole-brain data provide better MAE while using fewer features. The test results from using 256 and 1024 features of each table can be found in the supplement.

#### 4.3.1 Age estimation using T1WI

Table 4-1 and Table 4-2 show the age estimation results from using T1WI with different brain regions. The LBP-TOP histograms from CORT provide the most information for predicting age. This result might reflect the observation that gray matter density shows large changes in some cortical regions across the human lifespan [44]. Regions in SUB\_CORT and CERE provide little information for age estimation. Interestingly, considering only the GM and WM regions provides acceptable MAE. This result demonstrates that the change in brain morphology across the lifespan is so large that it

can affect the LBP-TOP distribution in the whole-brain region.



**Table 4-1 Mean absolute error (MAE) of brain age estimation by LBP-TOP and RBF SVR using T1WI data and 2048 features in various brain regions.**

atlas registration	image data		whole-head			GM segment		
	region	radius (voxels)	1	2	3	1	2	3
Linear	GM		8.75	<b>8.04</b>	8.96	7.01	6.78	<b>6.28</b>
		WM	8.99	<b>8.20</b>	8.77	7.01	7.11	<b>6.95</b>
	SUB_CORT	CORT	6.87	<b>6.85</b>	7.02	<u>5.89</u>	6.07	6.04
			7.10	7.18	<b>6.74</b>	7.89	<b>7.39</b>	7.40
		CERE	9.14	9.47	<b>8.91</b>	10.02	10.05	<b>9.59</b>
Non-linear	GM		8.70	<b>7.99</b>	8.94	6.77	6.73	<b>6.40</b>
		WM	8.87	<b>8.16</b>	8.82	7.33	7.23	<b>6.86</b>
	SUB_CORT	CORT	6.83	<b>6.63</b>	7.02	<b>5.95</b>	6.27	6.18
			7.40	7.21	<b>6.90</b>	7.85	7.29	<b>7.16</b>
		CERE	9.02	<b>8.74</b>	8.78	9.34	8.86	<b>8.45</b>

The bold values denote the best MAE in the image data of the row. The italic values denote the results using 354 features, which is the total number of features provided by the region. The underlined value shows the best MAE in the table.

and Table 4-4 show the results of combining the features from all of the brain regions. The best result from using the T1WI brain image is the G M segment data, where the MAE equals 5.62 years with 1 voxel as the radius. The scatter plots of the true and estimated age of each subject from this test are shown in Figure 4-5. As shown in Table 4-3, both the rough MNI regions and the total combined regions can enhance the age-predictive performance. MNI provided sufficiently good performance compared to using all of the combined regions. However, combining all of the regions consistently improved the results, especially while using the GM segment data. Despite the increased complexity

of the large feature number in the combined data, this result could imply that while most of the morphological changes across the lifespan are dependent, some in different brain regions are independent. Therefore, the combined information provides only a slight improvement to the overall performance.

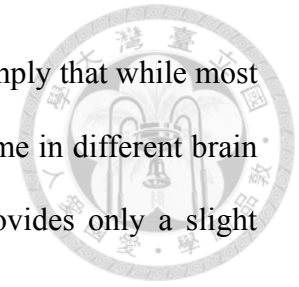


Table 4-2 Mean absolute error (MAE) of brain age estimation by LBP-TOP and RBF SVR using T1WI data in various brain regions.

atlas registration	image data			whole-head									GM segment										
	radius	region \ used features			1 voxel			2 voxels			3 voxels			1 voxel			2 voxels			3 voxels			
		256	1024	2048	256	1024	2048	256	1024	2048	256	1024	2048	256	1024	2048	256	1024	2048	256	1024	2048	
Linear	GM	8.65	8.75		8.22	<b>8.04</b>		9.01	8.96		6.73	7.01		6.98	6.78		6.37	<b>6.28</b>		6.37	<b>6.28</b>		
	WM	8.79	8.99		<b>8.09</b>	8.20		8.75	8.77		<b>6.92</b>	7.01		7.16	7.11		7.10	6.95		7.10	6.95		
	CORT	7.29	6.91	6.87	7.02	<b>6.76</b>	6.85	7.53	7.05	7.02	6.58	6.03	<b>5.89</b>	6.98	6.52	6.07	6.57	6.35	6.04	6.57	6.35	6.04	
	SUB_CORT	7.23	6.75	7.10	7.29	7.05	7.18	6.71	<b>6.56</b>	6.74	8.31	7.86	7.89	8.29	<b>7.38</b>	7.39	8.42	7.42	7.40	8.42	7.42	7.40	
	CERE	9.53	9.04	9.14	9.79	9.60	9.47	9.72	<b>8.87</b>	8.91	10.87	10.11	10.02	10.63	10.24	10.05	10.15	<b>9.67</b>	9.59	10.15	<b>9.67</b>	9.59	
		GM	8.45	8.70		8.01	<b>7.99</b>		8.94	8.94		6.64	6.77		6.77	6.73		6.47	<b>6.40</b>		6.47	<b>6.40</b>	
Non-linear	WM	8.71	8.87		<b>8.12</b>	8.16		8.75	8.82		7.23	7.33		7.26	7.23		6.96	<b>6.86</b>		6.96	<b>6.86</b>		
	CORT	7.06	6.87	6.83	6.93	<b>6.55</b>	6.63	7.27	6.93	7.02	6.44	6.07	<b>5.95</b>	7.07	6.55	6.27	7.37	6.44	6.18	7.37	6.44	6.18	
	SUB_CORT	7.86	7.41	7.40	7.59	7.21	7.21	6.96	<b>6.71</b>	6.90	8.46	7.88	7.85	8.13	7.47	7.29	8.48	7.21	<b>7.16</b>	8.48	7.21	<b>7.16</b>	
	CERE	10.34	8.97	9.02	8.95	<b>8.71</b>	8.74	8.99	8.80	8.78	10.71	9.64	9.34	9.50	8.92	8.86	9.14	8.61	<b>8.45</b>	9.14	8.61	<b>8.45</b>	
		GM	8.45	8.70		8.01	<b>7.99</b>		8.94	8.94		6.64	6.77		6.77	6.73		6.47	<b>6.40</b>		6.47	<b>6.40</b>	
		WM	8.71	8.87		<b>8.12</b>	8.16		8.75	8.82		7.23	7.33		7.26	7.23		6.96	<b>6.86</b>		6.96	<b>6.86</b>	

The bold values denote the best MAE in the image data of the row. The italic values denote the results using 354 features, which is the total number of features provided by the region. The underlined value shows the best MAE in the table.



**Table 4-3 Mean absolute error (MAE) of brain age estimation by LBP-TOP and RBF SVR using T1WI data and 2048 features in combined brain regions.**

atlas registration	image data		whole-head			GM segment		
	region	radius (voxels)	1	2	3	1	2	3
Linear	MNI		6.23	<b>5.95</b>	6.42	<b>5.68</b>	5.88	5.89
	GM+WM+CORT+SUB_CORT+CERE+MNI		6.27	<b>6.07</b>	6.24	<u>5.62</u>	5.64	5.73
Non-linear	MNI		6.34	<b>6.15</b>	6.37	<b>5.74</b>	6.14	5.77
	GM+WM+CORT+SUB_CORT+CERE+MNI		6.39	<b>6.10</b>	6.38	<b>5.63</b>	5.66	5.76

The bold values denote the best MAE in the image data of the row. The underlined value shows the best MAE in the table.

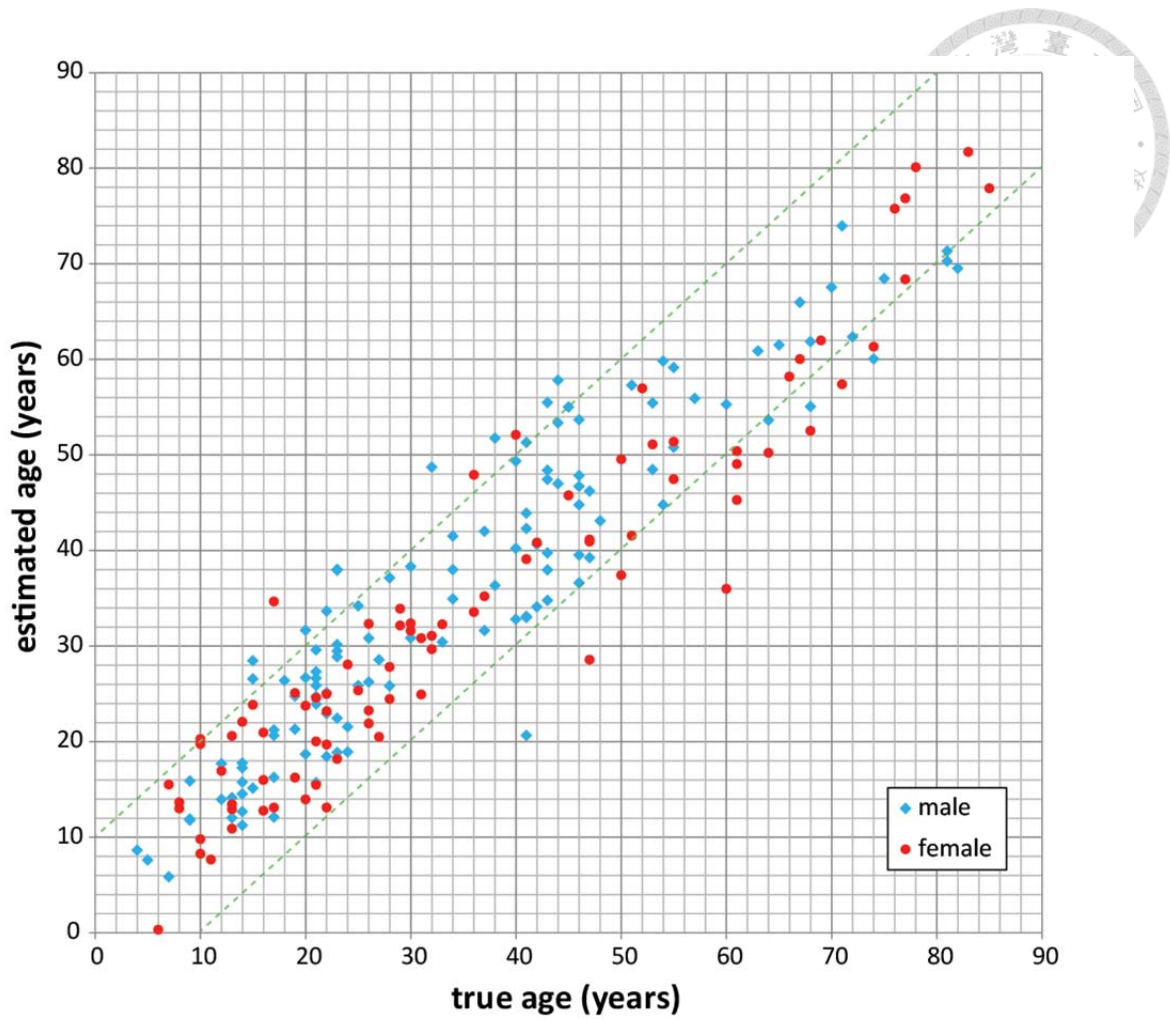


Table 4-4 Mean absolute error (MAE) of brain age estimation by LBP-TOP and RBF SVR using T1WI data in combin

	image data			whole-head									GM segment				
	region	radius	used features	1 voxel	2 voxels	3 voxels	1 voxel	2 voxels	3 voxels	1 voxel	2 voxels	3 voxels	1 voxel	2 voxels	3 voxels		
atlas registration				256	1024	2048	256	1024	2048	256	1024	2048	256	1024	2048		
Linear	MNI	GM+WM+CORT+SUB_CORT+CERE+MNI	7.00	6.17	6.23	6.51	6.03	<b>5.95</b>	7.13	6.47	6.42	6.38	6.02	<b>5.68</b>	6.49	5.98	5.88
			<b>6.15</b>	6.26	6.27	6.39	6.20	6.07	6.39	6.31	6.24	6.27	5.82	<b>5.62</b>	6.77	5.83	5.64
Non-linear	MNI	GM+WM+CORT+SUB_CORT+CERE+MNI	7.19	6.27	6.34	<b>6.24</b>	6.27	6.15	6.76	6.42	6.37	6.35	5.88	<b>5.74</b>	6.52	6.37	6.14
			<b>6.42</b>	6.42	6.39	6.67	6.11	6.10	6.47	6.42	6.38	6.39	5.89	<b>5.63</b>	6.90	5.96	5.66

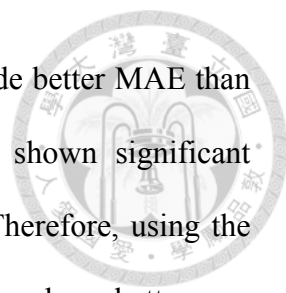
The bold values denote the best MAE in the image data of the row. The underlined value shows the best MAE in the table.





**Figure 4-5 Scatter plots of age estimation of all 204 subjects using GM segments.**

This figure shows the specific age estimation of each subject obtained from the best result in Table 4-3. The estimation models were trained with 2048 features using GM segments, linear atlas registration, a radius of 1 voxel, and combined brain regions (GM+WM+CORT+SUB\_CORT+CERE+MNI in Table 4-3). The mean absolute error is 5.62 years as shown in Table 4-3. The dashed green lines indicate the true age $\pm$ 10 years.

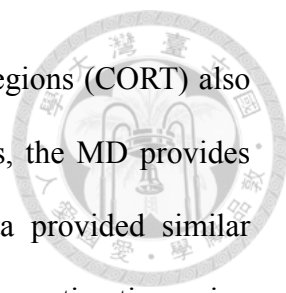


For the GM, WM, and cortical regions, the GM segment data provide better MAE than the whole-brain data. As mentioned before, Sowell et al. have shown significant changes in the gray matter density of some cortical regions [44]. Therefore, using the GM segment data can reflect this behavior more clearly and produce better age estimation results. On the other hand, for the sub-cortical (SUB\_CORT) and cerebellar (CERE) regions, the whole-brain data provide better performance. The best known aging-related change in those regions is an increase in CSF volume [44]. The GM segment data filter out the CSF segment and provides less information in the estimation of brain age.

The results in Table 4-1 and Table 4-3 show that LBP-TOP can extract discriminative information from either unnormalized GM segment data or unmodified and unnormalized whole-brain images to predict brain age. The GM segment data show better brain age estimation results, but using whole-brain image data can also provide good results. Brain segmentation is still a challenging task, and the usage is relatively limited in clinical practice, especially when dealing with abnormal brains [138]. Therefore, the characteristics of LBP-TOP can extract effective features from unmodified and unnormalized original whole-brain images, which could be very useful in clinical applications.

#### 4.3.2 *Age estimation using DTI data*

Table 4-5 and Table 4-6 show the results of DTI age estimation by combined brain segments. The results with separate brain regions are shown in Table 4-7. Similar to the



results of the T1WI approaches, the information from the cortical regions (CORT) also shows the best estimation accuracy. Among the DTI measurements, the MD provides the best performance. Being the major part of MD, the L1 data provided similar performance compared to MD. Fig. 6 shows the scatter plot of age estimation using those DTI measurements. FA, MD, and L1 show very similar age estimating patterns compared to the results from T1WI. The important and interesting result is that we can use the whole-brain DTI fiber direction (V1) to estimate brain age. Although V1 data produce inferior results, the data still provide a good estimation when the age is less than 60 years old. To the best of our knowledge, this could be the first report that subject age affects the neural fiber direction in the brain. LBP-TOP could be very useful to use the information embedded in DTI eigenvector direction.



**Table 4-5 Mean absolute error (MAE) of brain age estimation by LBP-TOP and RBF SVR using DTI data and 2048 features in combined brain regions.**

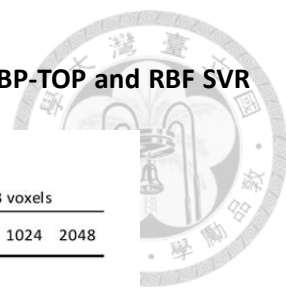
atlas registration	radius (voxels)	1	2	3
	DTI measurement			
linear	MD	7.35	<b>6.44</b>	6.53
	FA	7.20	7.37	<b>6.95</b>
	L1	7.43	6.70	<b>6.28</b>
	V1	9.72	8.47	<b>7.99</b>
non-linear	MD	6.83	<u>5.97</u>	6.32
	FA	6.94	6.84	<b>6.75</b>
	L1	6.74	6.08	<b>6.07</b>
	V1	8.74	<b>8.04</b>	8.07

The bold values denote the best MAE in the image data of the row. The underlined value shows the best MAE in the table.

**Table 4-6 Mean absolute error (MAE) of brain age estimation by LBP-TOP and RBF SVR using DTI data in combined brain regions.**

atlas registration	radius	1 voxel			2 voxels			3 voxels		
	used features	256	1024	2048	256	1024	2048	256	1024	2048
	DTI measurement									
linear	MD	7.91	7.53	7.35	6.99	<b>6.39</b>	6.44	7.17	6.65	6.53
	FA	8.04	7.16	7.20	8.17	7.37	7.37	7.28	<b>6.90</b>	6.95
	L1	7.69	7.37	7.43	6.92	6.68	6.70	6.49	<b>6.19</b>	6.28
	V1	10.23	9.76	9.72	8.68	8.19	8.47	8.84	8.00	<b>7.99</b>
non-linear	MD	7.10	6.86	6.83	6.42	<u>5.95</u>	5.97	6.73	6.21	6.32
	FA	7.30	6.87	6.94	7.03	6.82	6.84	6.69	<b>6.48</b>	6.75
	L1	7.45	6.82	6.74	6.44	<b>6.06</b>	6.08	6.21	5.92	6.07
	V1	9.07	8.86	8.74	9.11	8.30	<b>8.04</b>	9.17	8.39	8.07

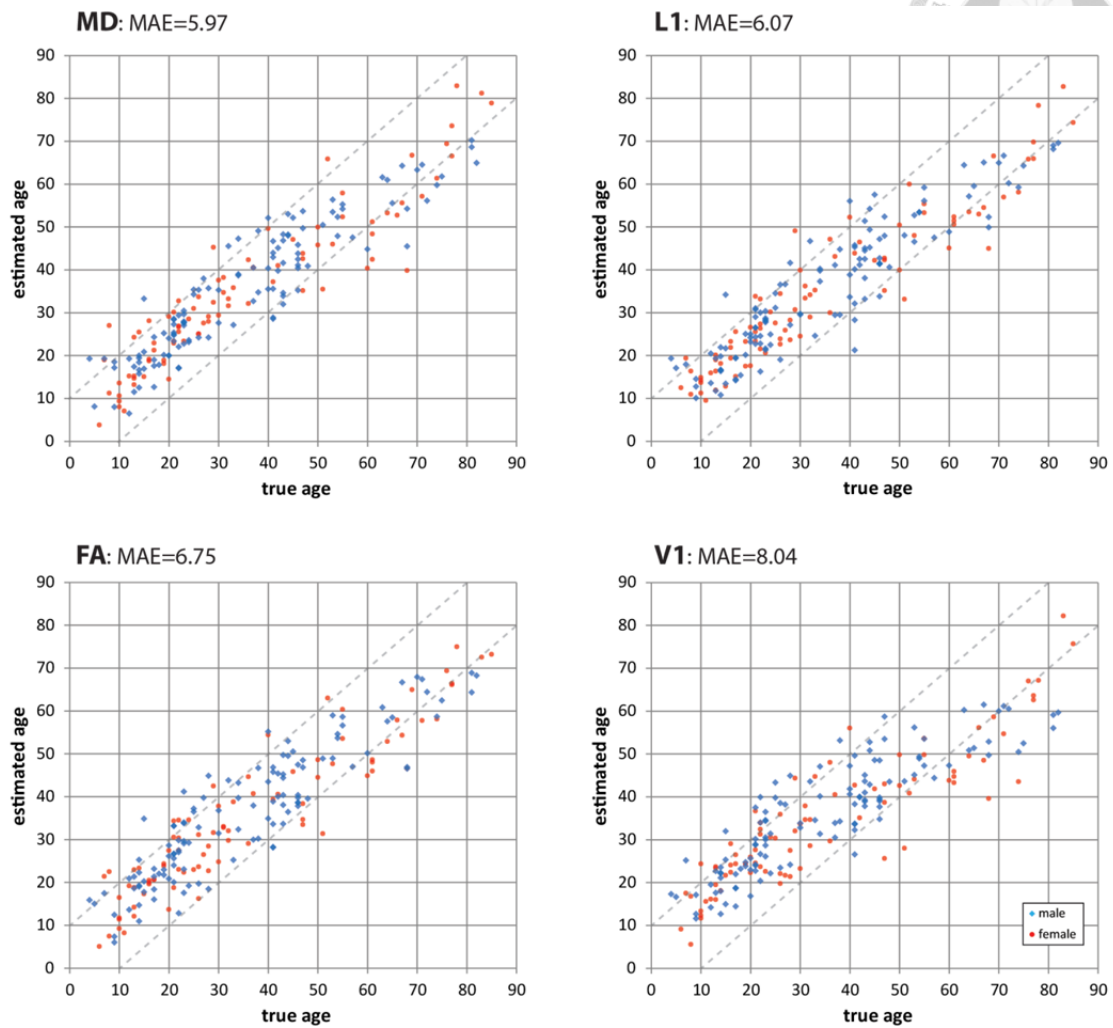
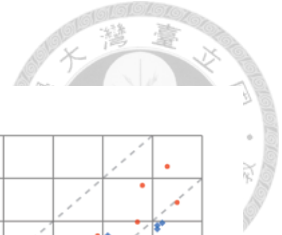
The bold values denote the best MAE in the image data of the row. The underlined value shows the best MAE in the table.



**Table 4-7 Mean absolute error (MAE) of brain age estimation using by LBP-TOP and RBF SVR DTI data in various brain regions.**

DTI measurement	atlas registration	radius		1 voxel			2 voxels			3 voxels			
		region	used features	256	1024	2048	256	1024	2048	256	1024	2048	
MD	Linear	GM		8.56			<b>7.41</b>			7.64			
			WM	8.27			<b>7.47</b>			7.57			
		SUB_CORT	CORT	7.38	7.25	7.35	7.13	6.63	6.59	6.82	<b>6.40</b>	6.52	
			CERE	9.15	8.44	8.55	8.10	7.83	8.00	8.32	<b>7.58</b>	7.75	
		MNI	11.86	11.19	11.15	11.29	10.65	10.78	10.33	10.31	<b>10.29</b>		
		Non-linear	GM		8.20			<b>7.31</b>			7.86		
	WM			8.01			7.66			<b>7.63</b>			
	SUB_CORT		CORT	6.88	6.77	6.84	6.36	6.31	<b>6.24</b>	6.71	6.46	6.60	
			CERE	8.97	7.73	7.95	7.88	<b>7.16</b>	7.31	8.33	7.23	7.48	
	MNI		11.25	10.67	10.60	10.05	<b>9.33</b>	9.45	9.73	9.66	9.66		
	FA		Linear	GM		7.92			<b>7.70</b>			8.32	
		WM			8.64			8.36			<b>8.28</b>		
SUB_CORT		CORT		8.28	<b>7.51</b>	7.59	9.04	8.11	7.95	8.21	8.02	7.79	
		CERE		8.86	7.89	8.13	8.35	7.88	7.85	7.66	<b>7.15</b>	7.16	
MNI		10.87		10.31	10.21	10.43	9.66	9.85	10.30	<b>9.42</b>	9.65		
Non-linear		GM			7.94			<b>7.80</b>			8.09		
			WM	8.39			8.45			<b>8.31</b>			
		SUB_CORT	CORT	8.40	7.42	<b>7.39</b>	8.04	7.52	7.41	8.17	7.58	7.40	
			CERE	8.30	8.04	8.16	7.26	7.05	7.25	7.54	<b>6.86</b>	7.14	
		MNI	10.52	10.02	10.04	9.73	<b>9.43</b>	9.54	9.83	9.75	9.67		
		L1	Linear	GM		8.05			<b>7.62</b>			7.82	
WM					8.38			<b>8.05</b>			8.22		
SUB_CORT	CORT			8.25	7.94	7.78	7.95	7.37	7.32	6.95	<b>6.76</b>	6.93	
	CERE			7.87	7.79	7.87	7.90	7.15	<b>7.10</b>	7.30	7.25	7.32	
MNI	12.72			10.82	10.94	10.58	9.83	9.95	10.32	<b>9.76</b>	9.91		
Non-linear	GM				7.66			<b>7.64</b>			7.71		
			WM	8.50			8.14			<b>7.87</b>			
	SUB_CORT		CORT	7.78	7.20	7.20	7.39	7.19	7.08	7.32	<b>6.79</b>	6.92	
			CERE	7.99	7.72	7.72	7.51	7.11	7.01	7.93	<b>7.00</b>	7.01	
	MNI		11.16	9.85	10.13	9.26	<b>9.16</b>	9.32	10.10	9.28	9.32		
	V1		Linear	GM		11.34	11.63		9.64	10.09		8.95	<b>8.70</b>
WM					11.32	11.40		10.26	10.21		8.91	<b>8.67</b>	
SUB_CORT		CORT		10.17	10.04	10.07	9.56	9.22	9.08	8.64	8.29	<b>8.22</b>	
		CERE		11.32	10.17	10.20	10.35	9.58	9.47	8.97	<b>8.82</b>	8.99	
MNI		12.09		12.28	12.14	<b>10.52</b>	10.96	11.07	11.85	10.96	10.96		
Non-linear		GM			10.03	9.52	9.51	9.02	8.47	8.37	8.65	8.05	<b>7.76</b>
			WM	9.98	10.13		8.86	9.36		8.75	<b>8.38</b>		
		SUB_CORT	CORT	11.28	11.62		10.42	10.35		<b>8.59</b>	<b>8.59</b>		
			CERE	8.75	8.37	8.24	9.16	8.11	7.94	9.10	8.15	<b>7.89</b>	
		MNI	11.15	10.07	10.24	10.61	9.63	9.39	10.05	<b>9.14</b>	9.31		
		MNI	12.72	12.16	11.66	11.30	<b>10.87</b>	11.12	11.50	10.68	10.57		
MNI		9.57	8.81	8.44	9.28	7.97	7.77	8.65	7.70	<b>7.55</b>			

The bold values denote the best MAE in the image data of the row. The underlined value shows the best MAE in the table.



**Figure 4-6 Scatter plots of age estimation of all 204 subjects using DTI measurements, FA, MD, L1(first eigenvalue), and V1(first eigenvector).**

This figure shows the specific age estimation of each subject from the best result in Table 4-5. The estimation models were trained with 2048 features using various DTI measurements, non-linear atlas registration, a radius of three voxels(FA and L1) or two voxels (MD and V1), and combined brain regions (GM+WM+CORT+SUB\_CORT+CERE+MNI in Table 4-3). The dashed lines indicate the true age $\pm$ 10 years.



### 4.3.3 *Atlas Registration*

In the estimation of age using T1WI data, linear registration and non-linear registration show the same performance. As shown in Table 4-1 and Table 4-3, using the linear or non-linear method to register the atlas information to the brain images did not make a big difference in the resulting MAE. The robustness of the atlas registration method could be an important characteristic for using LBP-TOP in clinical applications. It is easier to perform non-linear registration on healthy subjects, but abnormal brain morphology might interfere with the registration algorithm. Using LBP-TOP can produce sufficiently good results with linear atlas registration, extending its application range to many abnormal brain images.

Although the difference is not big, using non-linear atlas registration provide better performance for age estimation from the DTI data compared to the results using linear atlas registration (in Table 4-5, the difference in the MAE is less than 0.5 years on average). This could be because the lower resolution of the DTI data versus the T1WI data leads to a slight registration bias, which might affect the distribution of the patterns. Moreover, lower resolution introduces a larger partial volume effect when applying the atlas registration results. Therefore, low resolution data could be more sensitive to registration bias.





#### 4.3.4 *Learned Model*

Table 4-8 shows the proportion of features from the specific brain regions used in the top 256 ranked features of the age estimation models. Most of the age-related changes originate in GM. For the DTI measurements, changes in SUB\_CORT, CERE, and MNI played more important roles than in the GM segment data. In addition to morphological changes in brain structure, DTI can reveal microstructural changes in brain tissue. However, most age-related changes are also derived from the gray matter. The results of Table 4-8 also imply that although the data represent different measurements of brain tissue, they may be highly dependent. The changes in the DTI measurements might reflect changes in gray matter morphology or vice versa. Therefore, combining these data to build a mixed model does not greatly improve the total age estimation accuracy (data not shown). In studies of brain age estimation during brain maturation, only the information from the GM segment [107] provided similar accuracy compared to mixed models using signal intensity and DTI information [139].

**Table 4-8 The proportion of features from specific brain regions used in the top 256 features of the age estimation models.**

region	GM segment	FA	MD	L1	V1
GM	0.00%	0.00%	0.39%	0.00%	0.00%
WM	0.00%	0.00%	0.00%	0.00%	0.78%
MNI	5.86%	17.58%	11.33%	12.11%	6.64%
CORT	66.02%	31.25%	48.83%	41.80%	55.86%
SUB_CORT	12.89%	31.64%	20.70%	23.05%	17.19%
CERE	15.23%	19.53%	18.75%	23.05%	19.53%



### 4.3.5 Brain Maturation and Aging

Table 4-9 shows the published age estimation studies using MR brain images [107, 108, 111, 139, 140]. Our work is the only approach to extract information directly from unnormalized images. As shown in Table 4-9, greater error always results from wider age coverage. By providing the most extensive lifespan coverage, our methods produce high accuracy in the prediction of brain age. To estimate age using DTI data, we used the same database as the study by Mwangi et al. [111]. Without excluding any subjects in the database, we provided less estimation error by using LBP-TOP approaches. Moreover, because LBP-TOP extracted information from unmodified MR brain images, we can directly use the information of fiber spatial orientation (eigenvectors, V1) to predict brain age.

**Table 4-9 Comparison of age estimation studies using MR brain images.**


N	age range	$\Delta$ age	Target Period	MAE	Data Type	Method	Publication
885	3 ~ 20	17	Maturation	1.03	mixed	multiple modalities	Brown et al. 2012 [76]
394	5 ~ 18	13	Maturation	1.10	T1WI	PCA + RVR	Franke et al. 2012 [17]
550	19 ~ 86	67	Aging	4.98	GM segment	PCA + RVR	Franke et al. 2010 [18]
471	17 ~ 79	62	Aging	*6.50	T1WI	RVR	Ashburner et al. 2007 [77]
204	4 ~ 85	81	Lifespan	5.62	T1WI	LBP-TOP + SVR	proposed
204	4 ~ 85	81	Lifespan	5.97	DTI (MD)	LBP-TOP + SVR	proposed
**188	4 ~ 85	81	Lifespan	6.88	DTI (RD)	RVR	Mwangi et al. 2013 [21]

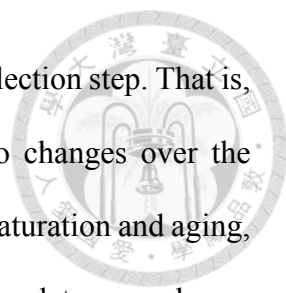
Only studies in which the number of test subjects is greater than 100 are listed. MAE: mean absolute error. \* denotes the root-mean-squared error. \*\* denotes that the study used the same database as the present work.

Despite the differences in training subject number and various methods, predicting age during the period of brain maturation (prior to 20 years of age) seems to be easier than estimating across the entire lifespan (Table 4-10). This trend is also apparent in previous studies [111, 139], where predicting the brain age of children is more accurate than that of elders. During childhood and adolescence, there seems to exist a programmed transformation of brain structure; therefore, we can easily trace the transformation either through MR anatomical images or brain neural tract changes using DTI. Although there are many reports of structural changes during brain aging [117, 141, 142], brain structure can be affected by genetic, epigenetic, and various environmental factors across the lifespan [143, 144]. Hence, predicting age throughout an aging period is more difficult.

**Table 4-10 Mean absolute error (MAE) of different age ranges of brain age estimation by LBP-TOP and RBF SVR using GM and DTI data (the results of specific subject are plotted in Figure 4-5 and Figure 4-6).**

	age range				
	< 15	16~30	31~45	46~60	> 61
<b>GM</b>	4.33	4.76	5.99	6.14	7.99
<b>MD</b>	5.51	4.48	5.62	5.76	10.55
<b>FA</b>	5.49	6.25	5.70	6.34	10.75
<b>L1</b>	5.48	4.93	6.15	5.52	9.78
<b>V1</b>	6.96	7.00	5.77	8.08	15.01

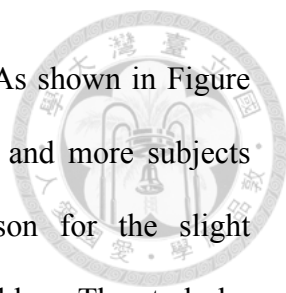




On the other hand, our age estimation approaches include a feature selection step. That is, our models were built using the features that are most relevant to changes over the lifespan. Therefore, if significantly different patterns exist between maturation and aging, only the features that are most relevant to changes across all ages in the data were chosen to build the fine RBF SVR models. The model might select too many features reflecting maturation-related brain changes to finely estimate brain age in elders. For better prediction, the use of different models to estimate maturation and aging should be considered in the future.

Our results using the DTI data in Figure 4-6 show an interesting pattern. Age estimation after 60 years old shows more error and is underestimated, especially in the models using the first eigenvectors (V1). The results could imply that there is no significant change in brain structure after the age of 60 in most of the participating subjects. The research of Sowell et al. [44] found a significant age effect in the dorsal frontal and parietal association cortices. Those effects consist of a dramatic decline in gray matter density between the ages of 7 and 60 with little or no decline thereafter. Their research could be evidence to confirm our estimated results. Our results imply that FA and MD might still change after the age of 60, but the fiber direction remains almost the same and provides no information for age estimation.

For better estimation accuracy, several aspects should be considered. First, as shown in Table 4-1, Table 4-3, and Table 4-5, using a different radius in LBP-TOP affects the result performance. There is no guideline for choosing the radius for specific resolution and data types. One must test and find the best radius for each application. Second, the



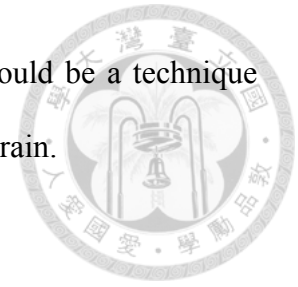
age of the training subjects should be balanced in each age range. As shown in Figure 4-1, the age distribution of the employed database is unbalanced, and more subjects were aged between 15~25 years old. This could be the reason for the slight over-estimation of age in children and under-estimation of age in elders. The study by Mwangi et al. used the same database and reported a similar trend of age estimation (Fig. 1 in [111]). Third, the training database should be large to properly train the estimation model. Studies of age estimation using facial images usually use thousands of images to train and build accurate models [145, 146]. While using more complex brain 3D volume data, more training subjects are needed to provide better accuracy.

## 4.4 Conclusion

In this study, we built an age estimation model as an imaging biomarker using LBP-TOP, a method that extracts information from unnormalized MR brain images, and support vector regression with T1WI or DTI measurements. The best estimated MAE using T1WI was 5.62 years was produced by LBP-TOP of the GM segment. We also showed the usability of LBP-TOP to extract age-related information from unnormalized DTI data. Among DTI measurements, MD provides the best MAE of 5.97 years.

This work shows that LBP-TOP can effectively extract discriminating information from unnormalized MR scale images or DTI measurements. In addition to scalar images, we also demonstrated that the information encoded in fiber orientation derived by DTI

eigenvectors can estimate brain age. This reveals that LBP-TOP could be a technique used for quantifying and comparing fiber information in the whole brain.





## Chapter 5

### ADHD classification using Local Binary Patterns

#### 5.1 Introduction

Attention-Deficit/Hyperactivity Disorder (ADHD) is a multifactorial and clinically heterogeneous disorder, which is highly prevalent in children worldwide. It is estimated that 5-10% of school-age children and 4% of adults suffer from ADHD [147]. The negative impact of ADHD on patients, their families, and society make ADHD a major public health problem [148]. However, an objective biological tool to diagnose ADHD is still unavailable. Foreseeing the importance, the organizers of the ADHD-200 Global

Competition have collected functional and anatomical ADHD MRI datasets of an unprecedented scale, which are accessible via the Internet ([http://fcon\\_1000.projects.nitrc.org/indi/adhd200/](http://fcon_1000.projects.nitrc.org/indi/adhd200/)). This work provides an important opportunity for researchers all over the world to study brain changes in ADHD subjects based on numerous brain MRI images.

Using the ADHD-200 database, we found that the brain morphological changes described by a 3D texture analysis can be used to distinguish children with ADHD from typically developing children (TDC). These structural image-based models demonstrated similar accuracy compared with our models based on rs-fMRI data. In the present study, we describe and analyze the 3D texture analysis method.

It is not easy to construct a classification rule to distinguish ADHD from TDC subjects. ADHD is a complex disorder with a composite etiology [149]. No simple existing indicators can be used to diagnose ADHD at present. Currently, the Diagnostic and Statistical Manual of Mental Disorders, 4th edition, text revision (DSM-IV-TR) is most often used for diagnostic criteria for ADHD. Some ADHD criteria are based on subjective descriptions by a child's parents or teachers and not on objective analysis tools. Recent research has demonstrated that using different versions of the DSM or disparate sources of collateral information can significantly affect the calculated prevalence of ADHD [147]. Moreover, both sex and age play important roles in the development of ADHD. These factors also increase the complexity of building a diagnostic tool [147]. All the aforementioned factors make it challenging to build an efficient classification model for ADHD.

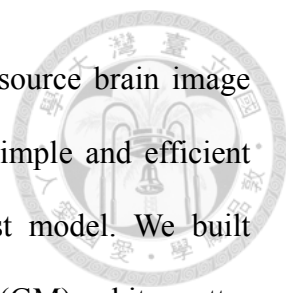


Additionally, approaches based on rs-fMRI data suffer from unstable echo planar imaging (EPI) and involve sophisticated data preprocessing steps. For these reasons, building a classification model based on rs-fMRI data from multiple research sites involves difficult manipulations of large data sets and is not efficient.

However, structural brain images are of high quality and are more stable with better resolution compared with rs-fMRI data. We hypothesized that structural brain images might contain more information from which to build a discriminative model. Although ADHD is not believed to result from morphological changes in the brain, several studies have shown that anatomical differences associated with ADHD can be found in MR images [147]. Large changes in volume and structural differences in the cerebral cortex have also been reported using MRI methodologies, such as anatomical MRI and diffusion tensor imaging [147]. Hence, we set forth to develop an ADHD classification method based on morphological changes. Notably, after using 3D texture descriptors to extract features from brain anatomical data, we found that morphological changes provided information that could discriminate ADHD from TDC subjects.

In this paper, to describe brain morphology, we introduce a feature extraction method based on texture point of view using the isotropic local binary patterns on three orthogonal planes (LBP-TOP). After extracting features using LBP-TOP, we trained a support vector machines (SVM) model and built an ADHD classification model based on the extracted features.

In the present study, we build an ADHD classification model using LBP-TOP features



and SVM. Different registration methods, LBP-TOP settings, and source brain image resolutions were utilized to test the properties of this method. A simple and efficient feature selection method was introduced to create a more robust model. We built classification models based on three basic brain tissues: gray matter (GM), white matter (WM), and CSF. Our results demonstrate that it is possible to build an ADHD classification model based on LBP-TOP features. We found that GM data provide the most salient information for discriminating ADHD from TDC subjects.

## **5.2 Materials and Methods**

### *5.2.1 Participants*

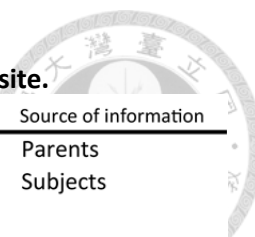
To best demonstrate the discriminative power of LBP-TOP, only male subjects were used to control for the known sex-based differences in ADHD subjects [147]. Male data from the Kennedy Krieger Institute (KKI), the NeuroIMAGE sample (NeuroIMAGE), the New York University Child Study Center (NYU), Oregon Health and Science University (OHSU), Peking University (Peking\_1, Peking\_2, and Peking\_3), and the University of Pittsburgh (Pittsburgh) were selected for analysis in this study. We ruled out using the dataset from Washington University because it was not in the test set of the ADHD-200 global competition and no ADHD subject in it. Five subjects (0010016, 0010027, 0010055, 0010098, and 0010127) in the NYU dataset were excluded because

no anatomical data existed for them. Subject 0010013 in the NYU dataset was also excluded because some of the brain in the anatomical image was cropped during the face removal process. ADHD hyperactive-type subjects were excluded due to the small number of such subjects in the dataset.

Therefore, the ADHD subjects in this study were of both the ADHD combined type and the ADHD inattentive type. A total of 436 male subjects (210 ADHD subjects and 226 TDC, mean age=12.12  $\pm$  2.95) were used in this study. The distributions of subjects by age and by type of ADHD are shown in Table 5-2 and Table 5-3. A list of all subjects can be found in Supp. Table 5-1. The detailed phenotype of each subject can be found on the website for the ADHD-200 global competition ([http://fcon\\_1000.projects.nitrc.org/indi/adhd200/](http://fcon_1000.projects.nitrc.org/indi/adhd200/)).

### 5.2.2 *Diagnostics of ADHD*

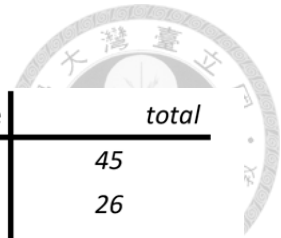
Table 5-1 shows a brief summary of the diagnosis criteria used by each site. The sites used different ADHD criteria, intellectual evaluations, and sources of collateral information.



**Table 5-1 A brief summary of the different diagnostic criteria used by each site.**

Site	ADHD criteria	Intelligence evaluation	Source of information
Kennedy Krieger Institute (KKI)	DICA-IV	WISC-IV	Parents
	DuPaul		Subjects
	CPRS-R		
	DSM-IV		
NeuroIMAGE sample (NeuroIMAGE)	KSADS-PL	WASI	Parents
	CPRS-LV		Subjects
New York University Child Study Center (NYU)	KSADS-PL	WASI	Parents
	CPRS-LV		Subjects
Oregon Health & Science University (OHSU)	KSADS-I	WISC-IV	Parents
	CPTRS-III		Teachers Subjects
Peking University (Peking)	C-DIS-IV	WISCC-R	Parents
	KSADS-PL		
University of Pittsburgh (Pittsburgh)	N/A	WASI	N/A

Abbreviations are as follows: **C-DIS-IV**, Computerized Diagnostic Interview Schedule IV; **CPRS-LV**, Conners' Parent Rating Scale-Revised, Long version; **CPRS-R**, Conners' Parent Rating Scale-Revised, Long Form; **CPTRS-III**, parent and teacher Connors' Rating Scale, Third Edition; **DuPaul**, DuPaul ADHD Rating Scale IV; **DSM-IV**, Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition; **KSADS-I**, Kiddie Schedule for Affective Disorders and Schizophrenia; **KSADS-PL**, Schedule of Affective Disorders and Schizophrenia for Children - Present and Lifetime Version; **WASI**, Wechsler Abbreviated Scale of Intelligence; **WISCC-R**, Intelligence Scale for Chinese Children-Revised; and **WISC-IV**, Wechsler Intelligence Scale for Children, Fourth Edition. Details can be found on the ADHD-200 global competition website ([http://fcon\\_1000.projects.nitrc.org/indi/adhd200/](http://fcon_1000.projects.nitrc.org/indi/adhd200/)).



**Table 5-2 Summary of the ADHD and TDC subjects used in this study.**

	TDC	ADHD Combined	ADHD Inattentive	<i>total</i>
KKI	34	7	4	45
NeuroIMAGE	11	14	1	26
NYU	46	59	29	134
OHSU	18	19	6	43
Peking_1	17	7	10	34
Peking_2	31	15	20	66
Peking_3	23	7	12	42
Pittsburgh	46	0	0	46
<i>total</i>	226	128	82	436

Abbreviations are as follows: **TDC**, typically developing children; **KKI**, the Kennedy Krieger Institute; **NeuroIMAGE**, the NeuroIMAGE sample; **NYU**, the New York University Child Study Center; **OHSU**, Oregon Health and Science University; **Peking\_1**, **Peking\_2**, and **Peking\_3**, Peking University; and **Pittsburgh**, the University of Pittsburgh.

**Table 5-3 The age distribution of subjects used in this study.**

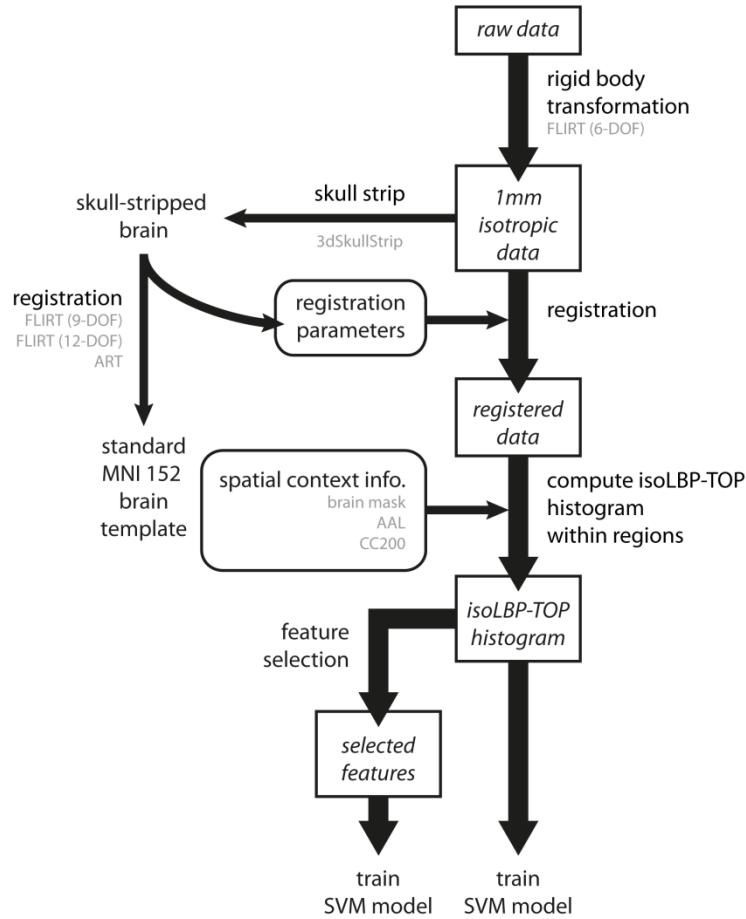
	TDC	ADHD Combined	ADHD Inattentive	<i>total</i>
KKI	10.32 ±1.37	9.9 ±1.52	11.46 ±1.19	10.36 ±1.4
NeuroIMAGE	17.38 ±2.27	17.04 ±2.19	18.8 ±0	17.25 ±2.16
NYU	12.21 ±3.16	10.93 ±2.58	12.48 ±2.78	11.71 ±2.89
OHSU	8.94 ±1.43	8.76 ±1.09	8.9 ±1.31	8.85 ±1.24
Peking_1	11.8 ±1.32	10.7 ±1.15	11.88 ±2.82	11.6 ±1.86
Peking_2	11.66 ±1.81	12.08 ±2.07	12.95 ±1.44	12.15 ±1.83
Peking_3	13.21 ±0.95	12.33 ±1.06	13.84 ±1.21	13.24 ±1.14
Pittsburgh	14.57 ±2.87	N/A	N/A	14.57 ±2.87
<i>total</i>	12.39 ±2.99	11.42 ±3.05	12.49 ±2.5	12.12 ±2.95

**TDC**, typically developing children.

### 5.2.3 *Data preprocessing*

An overview of the data analysis procedure is shown in Figure 5-1. The details of each step are described below.





**Figure 5-1 Overview of the data analysis procedure.**

Preprocessing involves three steps. First, all raw data are transformed into a 1 mm isotropic volume using the rigid body transformation as performed by FLIRT with 6 degrees of freedom (6-DOF). Second, brain images are registered to standard MNI152 space by linear (FLIRT with 9-DOF and FLIRT with 12-DOF) and non-linear (ART) registration methods. To achieve better registration results, the registration parameters were obtained by transforming the skull-stripped brains to the standard MNI152 brain template. Third, we computed the LBP-TOP histograms based on the registered images with various spatial context information (i.e., brain mask, AAL, and CC200). Following these steps, classification models can be trained by directly using the resulting histogram or by using a subset of data after applying the feature selection algorithm. ( The major process flow is denoted by the thick line. The minor process flow is denoted by the thin line. The square boxes are the major steps showing how to extract features from raw data. The round boxes are the parameters or information needed for the process flow. Texts in gray color indicate the methods or the subtypes the process step used. )

### Registration methods



Linear registrations with 9 degree of freedom (9-DOF) and 12-DOF were performed using the linear multimodality registration method developed by Oxford FSL FLIRT [123, 130]. All images were transformed to standard MNI152 space by FLIRT with 6-DOF (rigid-body transformation). The results of FLIRT with 6-DOF were then linearly transformed by FLIRT with 9-DOF (rigid-body + independent scaling). The results of FLIRT with 9-DOF were also linearly transformed by FLIRT with 12-DOF (rigid-body + scales + skews). Non-linear normalization procedures were performed using the automated registration tool (ART). ART was developed by Ardekani et al. [147] and can be downloaded from <http://www.nitrc.org/projects/art/>. Klein and colleagues demonstrated that ART provides better efficiency and consistency than other non-linear registration methods [147].

### 3D skull stripping

To obtain better registration results, skull stripping was performed prior to using the registration algorithm, using the 3dSkullStrip algorithm [125] developed by AFNI [150]. 3dSkullStrip has proven to be a relatively robust skull-stripping algorithm [147]. However, it is not a perfect tool. Incomplete skull stripping can result in a loss of information from some brain regions. Hence, we applied the transformation parameters for skull-stripped brains to the original whole-brain images to create the input for the LBP-TOP algorithm.



### Spatial context information and brain parcellations



To examine different spatial context information, we performed three separate parcellations in this study.

First, a simple brain mask in MNI152 space, as provided by FSL, was used to compute the total histogram of the whole-brain volume. Second, to introduce spatial context based on brain anatomical information, the widely used automated anatomical labeling (AAL) template with 116 regions was used [151]. Finally, we also used an atlas derived from functionally parcellating the resting state data [152]. A 200 ROI version with 190 regions of spatially constrained parcellation (CC200) was used to introduce the spatial context information based on rs-fMRI data. The CC200 functional parcellation template made for the competition was kindly provided by Cameron Craddock. Details of the construction of CC200 have been previously published [152] and can also be found on the Athena preprocessing strategies page of the ADHD-200 preprocessed data website: <http://www.nitrc.org/plugins/mwiki/index.php/neurobureau:AthenaPipeline>.

### Computation of LBP-TOP

The LBP-TOP algorithm was implemented using Java to build the LBP-TOP map from the structural image. All resulting LBP-TOP histograms were mapped for the detection of uniform patterns. Preliminary testing (not shown) demonstrated that only the LBP-TOP with eight neighbors provided sufficient information to classify ADHD

within a reasonable processing time. Therefore, only tests with eight neighbors are shown here.



### Classifiers

A k nearest neighbor classifier (KNN, K=1) was used to show the baseline of the discriminative power of LBP-TOP. Moreover, an efficient and widely used classifier, SVM, was used in this work [97]. SVM maps training data into high-dimensional feature space to find the separating hyperplane with the maximal margin. Due to the large feature size of LBP-TOP results, we used linear SVM for greater efficiency. LIBLINEAR [100] was chosen for use because of its optimization for linear SVM.

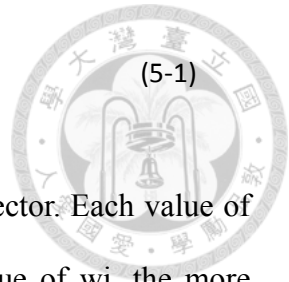
### Feature selection

After introducing special context information, the LBP-TOP histogram bins become an over-completed feature set. To build a more efficient and robust classification model, a feature selection method is needed. Moreover, by only selecting the most important features, we can combine features from various points of view. For example, we can combine features from different LBP-TOP results based on dissimilar radii.

Feature selection based on the linear SVM has proven to be efficient and useful for gene selection, document classification and many other applications [147].

For any test subject  $x$ , the decision function of linear SVM is

$$P(x) = \text{sign}(w^T x + b)$$



where  $x$  is the feature vector,  $b$  is a constant, and  $w$  is the weight vector. Each value of  $w$  denotes the weight of each feature. The larger the absolute value of  $w_j$ , the more important the  $j$ th feature is in deciding the result.

After training a linear SVM model, the  $w$  in (2-1) can be used as a relative importance index. Therefore, we can build a simpler model using the top  $n$  important features.

For combining features from different point of views, we first trained a linear SVM model using feature groups and ranked features by the absolute weights of the model. Only half of the features remained. Then, we combine these features with the features from a second feature group and trained another linear SVM model. Similarly, only half of the features were chosen to be merged into the next feature group. Using this iterative procedure, we combined various feature groups and found the most important features among these feature groups. Algorithm 5-1 shows steps of this iteration. Given a set of  $N$  subjects and  $K$  different feature groups, for each training dataset of our 10-fold cross-validation, we use Algorithm 5-1 to select and combine the most important features.



Algorithm 5-1 **The algorithm of feature selection.**

**Input :** Training dataset  $D = \left\{ \left\{ \mathbf{x}_{k,n}, y_n \right\}_{n=1}^N \right\}_{k=1}^K$

,  $\mathbf{x}_k$  is different feature set based on various setting

**Output :** Training dataset with selected features  $S = \left\{ \mathbf{s}_n, y_n \right\}_{n=1}^N$

For  $k = 1, \dots, K$

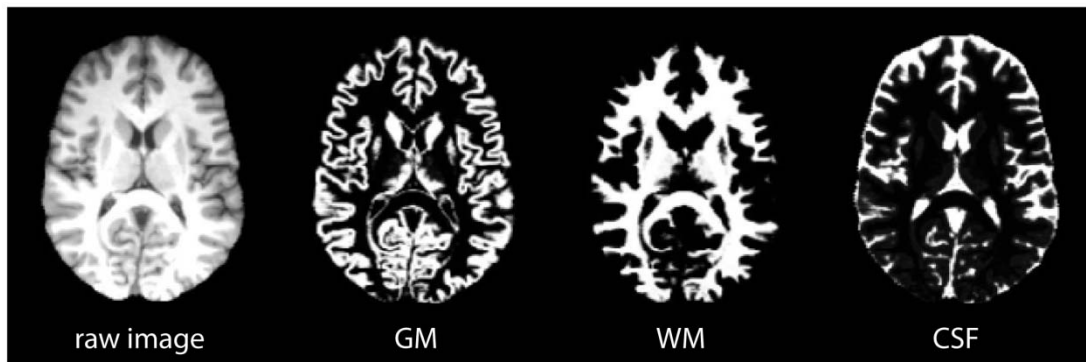
1. for each subject  $n$ , add all features of  $\mathbf{x}_{k,n}$  to  $\mathbf{s}_n$ .
2. use grid search with 10-fold cross-validation to find the best penalty parameter of linear-SVM based on  $S$ .
3. train a linear-SVM model based on  $S$  using the best penalty parameter.
4. sort the features of  $\mathbf{s}_n$  based on the absolute weights of the linear-SVM model.
- \*5. for each subject  $n$ , drop the last half features of  $\mathbf{s}_n$ .

Loop

\*Due to the small number of features revealed when analyzing the whole-brain region, we simply combined all the features and do not drop the last half of them.

### Brain segmentation

FSL's automated segmentation toolbox (FAST) was used to segment raw brain images into gray matter (GM), white matter (WM) and CSF [123, 124]. Figure 5-2 shows an example of a resulting probability map. The three tissue probability maps were analyzed following the same procedure described in Figure 5-1.

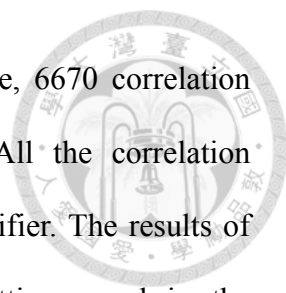


**Figure 5-2** Examples of brain probability maps based on gray matter (GM), white matter (WM), and CSF.

### Reference models based on rs-fMRI features

To compare the results of discriminative models based on rs-fMRI data, we used a simple and easily repeatable approach. Briefly, for the preprocessing of rs-fMRI data, we used the extracted timecourses from the Athena preprocessed data, which can be download from the ADHD-200 Preprocessed Data website. Details of the specific preprocessing steps can be found on the website.

The timecourses of the AAL and CC200 parcellations used in LBP-TOP study were chosen for comparison. The extracted timecourses files, ADHD200\_AAL\_TCs\_filtfix.tar.gz and ADHD200\_CC200\_TCs\_filtfix.tar.gz, can be found on the ADHD-200 Preprocessed Data website. The correlation coefficients between each pair of regions were computed based on their extracted timecourses. For



example, there are 116 regions in the AAL parcellation. Therefore, 6670 correlation coefficients can be computed based on the 6670 ROI pairs. All the correlation coefficients were used as features for the linear SVM ADHD classifier. The results of each model were validated using the same cross-validation settings used in the LBP-TOP studies. As described on the ADHD-200 Preprocessed Data website, the nuisance variance for the extracted time series of each region was removed, with or without use of a band-pass filter (0.009Hz ~ 0.08Hz), and blurred with a 6-mm FWHM Gaussian filter. Both time series, with or without filtering by a band-pass filter, were tested.

### Evaluation

All tests in this study were evaluated by 10-fold cross-validation. We randomly partitioned the 436 subjects into 10 subgroups. For each step of cross-validation, one subgroup was used as a test data set, and the remaining nine subgroups were pooled as a training data set. After 10 cross-validations, the test results of all 10 subgroups were combined to build the accuracy of the estimation of each model. To facilitate comparison of the results, the same 10-fold cross-validation set was used in all evaluations.

We used grid searching to find the best penalty parameter C for linear SVM for each training dataset. That is, another 10-fold cross-validation was applied to each training dataset with several candidate values of C, and we chose the parameter C that led to the highest accuracy.

While performing feature selection, the assignment of optimal feature weights can be achieved when the optimal value of  $C$  is chosen during each round of cross-validation. After that, we evaluated the effect of feature number using each testing dataset. Then, we combined the results of 10 test dataset to build the accuracy of different feature numbers.

### Statistical tests

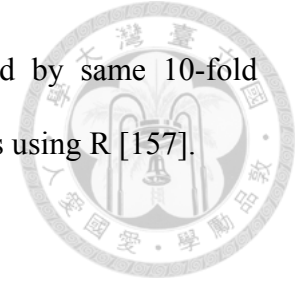
To show the classifier has learned a structure in the data, we compute the  $p$ -value against the null distribution using permutation tests [153, 154]. The null hypothesis of permutation test is that the labels are independent of the features. Therefore, one can learn almost same accuracy using random labeled data set. By randomly permuting the labels of the data set, permutation tests can measure how likely the observed accuracy is learned by chance. The permutation-based  $p$ -value is defined by

$$p = \frac{|\{D' \in \hat{D}: e(c, D') \leq e(c, D)\}| + 1}{k + 1} \quad (5-2)$$

where  $D$  is the original labeled data,  $e(c, D)$  denotes the error of classifier  $c$  learned from  $D$ , and  $\hat{D}$  is a set of  $k$  randomized versions  $D'$  of  $D$  [153]. In this work,  $e(c, D)$  was estimated by same 10-fold cross-validation with other tests. One hundred randomized sets of each test were used to estimate the  $p$ -values ( $k = 100$ ).

To compare different approaches of this work, McNemar's tests were applied to compute  $p$ -values between two approaches [155, 156]. While comparing two different

approaches, confusion matrices of each approach were estimated by same 10-fold cross-validations. Then we compute the  $p$ -values of McNemar's tests using R [157].



## 5.3 Results

### 5.3.1 *LBP-TOP*

Table 5-4 shows the 10-fold cross-validation results for different radii (1 mm, 2 mm, and 3 mm) for the LBP-TOP, various parcellations, linear registrations, and non-linear registrations, respectively. Table 5-4 (a) shows the baseline accuracy which LBP-TOP can provide with the simple 1NN classifier. Comparing the results of Table 5-4 (a) and Table 5-4 (b), we can find the linear-SVM classifiers can provide better accuracy than 1NN classifiers. Moreover, some of the properties changed while using different classifiers. The LBP-TOP with a radius equal to 3 mm provided better accuracy than the LBP-TOP for the other two radii in most cases while using linear-SVM classifiers. The same properties cannot be found while using 1NN approaches. However, there are nonsignificant between different radii in McNemar's test (Table 5-7). As expected, brain data with ART non-linear registration showed the highest accuracy in almost all cases, especially while using 1NN as classifiers. Notably, using linearly registered brain data did not greatly reduce accuracy. After apply McNemar's test, there is no significant difference between registration methods in any cases with linear-SVM classifiers. And only few



cases show significant difference between registration methods while using 1NN approaches (Table 5-8).



Although the resulting feature sizes varied widely (from 177 to 33630 features), accuracy across disparate parcellations was not greatly affected. Models using a histogram computed from the whole-brain region had higher accuracies than models based on other parcellations. Only considering the results of the AAL and CC200 parcellations, the CC200 showed better results most often. This finding may be the result of the greater number of utilized features or the greater number of homogeneous areas in the CC200 parcellation.

The results of reference models based on rs-fMRI features are shown in Table 5-4. These data indicate that simple approaches to analyzing rs-fMRI data do not discriminate as well as models based on structural information. The McNemar's test between structural features and rs-fMRI features also show significant difference in most cases (Table 3). Based on our experience in the ADHD-200 Global Competition, different preprocessing settings can affect the resulting accuracy. Moreover, combining the results of different rs-fMRI approaches can provide better discriminative power. The results of these simple approaches can be viewed as the baseline of discriminative power that rs-fMRI data can achieve.



**Table 5-4 (a, b) The ADHD-TDC classification accuracy of models based on LBP-TOP features with different registration methods, parcellation, and radius of LBP-TOP, using 1NN and linear-SVM classifiers alternatively.**

(a) 1NN models based on LBP-TOP features

parcellations	radius	features	DOF9	DOF12	ART
brain mask	R1	177	0.6009	<b>0.6376</b>	0.6009
	R2	177	0.6261	0.6399	<b>0.6537</b>
	R3	177	0.6239	<b>0.6422</b>	0.6353
AAL	R1	20532	0.5986	0.6009	<b>0.6422</b>
	R2	20532	0.5826	0.6032	<b>0.6353</b>
	R3	20532	0.5894	0.5826	<b>0.6376</b>
CC200	R1	33630	0.5826	0.6124	<b>0.6491</b>
	R2	33630	0.5917	0.6055	<b>0.6445</b>
	R3	33630	0.5940	0.5894	<b>0.6124</b>

(b) linear-SVM models based on LBP-TOP features

parcellations	radius	features	DOF9	DOF12	ART
brain mask	R1	177	0.6422	0.6376	<b>0.6514</b>
	R2	177	0.6606	0.6468	<b>0.6651</b>
	R3	177	0.6514	0.6399	<b>0.6583</b>
AAL	R1	20532	0.6239	0.6216	<b>0.6537</b>
	R2	20532	0.5986	<b>0.6261</b>	0.6239
	R3	20532	0.6284	0.6399	<b>0.6537</b>
CC200	R1	33630	0.6445	<b>0.6491</b>	0.6239
	R2	33630	0.6376	0.6491	<b>0.6537</b>
	R3	33630	0.6583	0.6560	<b>0.6697</b>

(c) 1NN models based on simple rs-fMRI features

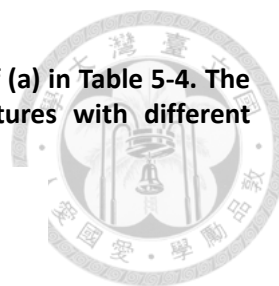
parcellations	features	non-BF	BF
AAL	6697	0.5459	<b>0.5688</b>
CC200	17955	<b>0.5550</b>	0.5528

(d) linear-SVM models based on simple rs-fMRI features

parcellations	features	non-BF	BF
AAL	6697	0.5665	<b>0.5734</b>
CC200	17955	<b>0.5803</b>	0.5596



The highest accuracy for each parcellation is denoted by the bold number. Abbreviations are as follows: **R1**, **R2**, and **R3**, LBP-TOP radius in mm; **DOF9**, and **DOF12**, linear registration with 9, and 12 degree of freedom, respectively; **ART**, non-linear registration performed by Automated Registration Tool; **AAL**, automated anatomical labeling template; and **CC200**, spatially constrained parcellation based on rs-fMRI. **(c, d) The ADHD-TDC classification accuracy of models based on simple rs-fMRI features, using 1NN and linear-SVM classifiers alternatively.** Abbreviations are as follows: **BF**, rs-fMRI data filtered by a bandpass filter (0.009Hz ~ 0.08Hz); and **non-BF**, rs-fMRI data not filtered by a bandpass filter. The sensitivity, specificity, and areas under the ROC curve (AUC) of this table can be found in Table 5-5 and Table 5-6.



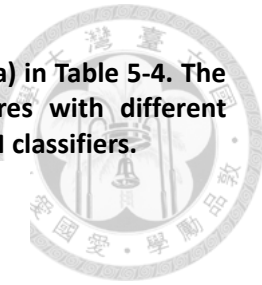
**Table 5-5 Sensitivity, specificity, and areas under the ROC curve (AUC) of (a) in Table 5-4. The ADHD-TDC classification accuracy of models based on LBP-TOP features with different registration methods, parcellation, and radius of LBP-TOP, using 1NN.**

<b>Sensitivity</b>					
parcellations	radius	features	DOF9	DOF12	ART
brain mask	R1	177	0.6504	0.6681	0.6504
	R2	177	0.6726	0.6770	0.6814
	R3	177	0.6637	0.6681	0.6549
AAL	R1	20532	0.6770	0.6991	0.6947
	R2	20532	0.6549	0.6903	0.6681
	R3	20532	0.6991	0.6726	0.6681
SCP200	R1	33630	0.6283	0.6681	0.6681
	R2	33630	0.6504	0.6681	0.6637
	R3	33630	0.6637	0.6504	0.6283

<b>Specificity</b>					
parcellations	radius	features	DOF9	DOF12	ART
brain mask	R1	177	0.5476	0.6048	0.5476
	R2	177	0.5762	0.6000	0.6238
	R3	177	0.5810	0.6143	0.6143
AAL	R1	20532	0.5143	0.4952	0.5857
	R2	20532	0.5048	0.5095	0.6000
	R3	20532	0.4714	0.4857	0.6048
SCP200	R1	33630	0.5333	0.5524	0.6286
	R2	33630	0.5286	0.5381	0.6238
	R3	33630	0.5190	0.5238	0.5952

<b>AUC</b>					
parcellations	radius	features	DOF9	DOF12	ART
brain mask	R1	177	0.5990	0.6365	0.5990
	R2	177	0.6244	0.6385	0.6526
	R3	177	0.6223	0.6412	0.6346
AAL	R1	20532	0.5956	0.5972	0.6402
	R2	20532	0.5798	0.5999	0.6341
	R3	20532	0.5853	0.5791	0.6365
SCP200	R1	33630	0.5808	0.6103	0.6484
	R2	33630	0.5895	0.6031	0.6438
	R3	33630	0.5914	0.5871	0.6118

Abbreviations are as follows: **R1**, **R2**, and **R3**, LBP-TOP radius in mm; **DOF9**, and **DOF12**, linear registration with 9, and 12 degree of freedom, respectively; **ART**, non-linear registration performed by Automated Registration Tool; **AAL**, automated anatomical labeling template; and **CC200**, spatially constrained parcellation based on rs-fMRI. The AUC was calculated using [106] and the positive subject is the TDC subjects.



**Table 5-6 Sensitivity, specificity, and areas under the ROC curve (AUC) of (a) in Table 5-4. The ADHD-TDC classification accuracy of models based on LBP-TOP features with different registration methods, parcellation, and radius of LBP-TOP, using linear-SVM classifiers.**

Sensitivity					
parcellations	radius	features	DOF9	DOF12	ART
brain mask	R1	177	0.6637	0.6726	0.6681
	R2	177	0.6903	0.6593	0.6947
	R3	177	0.6770	0.6504	0.6770
AAL	R1	20532	0.6637	0.6504	0.6593
	R2	20532	0.6062	0.6593	0.6327
	R3	20532	0.6637	0.6681	0.6637
SCP200	R1	33630	0.6637	0.6593	0.6062
	R2	33630	0.6593	0.6637	0.6681
	R3	33630	0.6504	0.6504	0.6814

Specificity					
parcellations	radius	features	DOF9	DOF12	ART
brain mask	R1	177	0.6190	0.6000	0.6333
	R2	177	0.6286	0.6333	0.6333
	R3	177	0.6238	0.6286	0.6381
AAL	R1	20532	0.5810	0.5905	0.6476
	R2	20532	0.5905	0.5905	0.6143
	R3	20532	0.5905	0.6095	0.6429
SCP200	R1	33630	0.6238	0.6381	0.6429
	R2	33630	0.6143	0.6333	0.6381
	R3	33630	0.6667	0.6619	0.6571

AUC					
parcellations	radius	features	DOF9	DOF12	ART
brain mask	R1	177	0.6414	0.6363	0.6507
	R2	177	0.6594	0.6463	0.6640
	R3	177	0.6504	0.6395	0.6575
AAL	R1	20532	0.6223	0.6205	0.6535
	R2	20532	0.5983	0.6249	0.6235
	R3	20532	0.6271	0.6388	0.6533
SCP200	R1	33630	0.6438	0.6487	0.6245
	R2	33630	0.6368	0.6485	0.6531
	R3	33630	0.6586	0.6562	0.6693

Abbreviations are as follows: **R1**, **R2**, and **R3**, LBP-TOP radius in mm; **DOF9**, and **DOF12**, linear registration with 9, and 12 degree of freedom, respectively; **ART**, non-linear registration performed by Automated Registration Tool; **AAL**, automated anatomical labeling template; and **CC200**, spatially constrained parcellation based on rs-fMRI. The AUC was calculated using [106] and the positive subject is the TDC subjects.

**Table 5-7  $p$ -values of McNemar's test of 1NN models and linear-SVM models based on LBP-TOP features between different radii (R1, R2, R3).**



(a) McNemar's test of 1NN models based on LBP-TOP features

parcellations	radius	DOF9	DOF12	ART
brain mask	R1 and R3	0.4403	0.9247	0.2699
	R2 and R3	<u>1.0000</u>	1.0000	<u>0.4606</u>
	R1 and R2	0.2891	1.0000	0.0602
AAL	R1 and R3	<u>0.7664</u>	<u>0.4966</u>	<u>0.9331</u>
	R2 and R3	0.7794	<u>0.2225</u>	1.0000
	R1 and R2	<u>0.4941</u>	1.0000	<u>0.8557</u>
CC200	R1 and R3	0.6935	<u>0.3583</u>	<u>0.1883</u>
	R2 and R3	1.0000	<u>0.3914</u>	<u>0.1359</u>
	R1 and R2	0.7373	<u>0.8041</u>	<u>0.9247</u>

(b) McNemar's test of linear-SVM models based on LBP-TOP features

parcellations	radius	DOF9	DOF12	ART
brain mask	R1 and R3	0.7237	1.0000	0.8423
	R2 and R3	<u>0.6511</u>	<u>0.7277</u>	<u>0.8150</u>
	R1 and R2	0.3816	0.7119	0.5940
AAL	R1 and R3	0.8897	0.3222	1.0000
	R2 and R3	0.0425	0.3613	0.0485
	R1 and R2	<u>0.1531</u>	0.8711	<u>0.1366</u>
CC200	R1 and R3	0.4962	0.7874	<b>0.0272</b>
	R2 and R3	0.1237	0.7194	0.2812
	R1 and R2	<u>0.7423</u>	1.0000	0.1182

The  $p$ -values without under lines denote the accuracies of R3 are bigger than R2, the accuracies of R3 are bigger than R1, and the accuracies of R2 are bigger than R2. The underlined  $p$ -values show the inverse relationship. Bold  $p$ -values correspond to significant results ( $p$ -value < 0.05). **DOF9**, and **DOF12**, linear registration with 9, and 12 degree of freedom, respectively; **ART**, non-linear registration performed by Automated Registration Tool; **AAL**, automated anatomical labeling template; and **CC200**, spatially constrained parcellation based on rs-fMRI.

**Table 5-8 p-values of McNemar’s test of 1NN models and linear-SVM models based on LBP-TOP features between different registrations.**



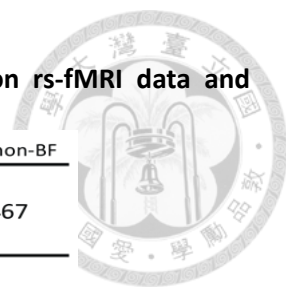
(a) McNemar’s test of 1NN models based on LBP-TOP features

parcellations	radius	DOF9 and ART	DOF12 and ART	DOF9 and DOF12
brain mask	R1	1.0000	<u>0.2415</u>	0.1297
	R2	0.3309	0.6508	0.6060
	R3	0.7139	<u>0.8533</u>	0.4606
AAL	R1	0.1268	0.1479	1.0000
	R2	0.0564	0.2392	0.2976
	R3	0.0760	<b>0.0373</b>	<u>0.8097</u>
CC200	R1	<b>0.0245</b>	0.2176	0.1715
	R2	<b>0.0437</b>	0.1254	0.5186
	R3	0.5329	0.3951	<u>0.8918</u>

(b) McNemar’s test of linear-SVM models based on LBP-TOP features

parcellations	radius	DOF9 and ART	DOF12 and ART	DOF9 and DOF12
brain mask	R1	0.7806	0.6272	<u>0.9049</u>
	R2	0.9187	0.5006	<u>0.4795</u>
	R3	0.8220	0.4220	<u>0.5827</u>
AAL	R1	0.2460	0.1936	<u>1.0000</u>
	R2	0.2891	<u>1.0000</u>	0.1198
	R3	0.2837	0.5713	0.5108
CC200	R1	<u>0.4068</u>	<u>0.2945</u>	0.8828
	R2	0.4825	0.9035	0.4414
	R3	0.6301	0.5383	<u>1.0000</u>

The  $p$ -values without under lines denote the accuracies of ART are bigger than DOF9, the accuracies of ART are bigger than DOF12, and the accuracies of DOF12 are bigger than DOF9. The underlined  $p$ -values show the inverse relationship. Bold  $p$ -values correspond to significant results ( $p$ -value < 0.05). **DOF9**, and **DOF12**, linear registration with 9, and 12 degree of freedom, respectively; **ART**, non-linear registration performed by Automated Registration Tool; **AAL**, automated anatomical labeling template; and **CC200**, spatially constrained parcellation based on rs-fMRI.



**Table 5-9  $p$ -values of McNemar’s test of linear-SVM models based on rs-fMRI data and LBP-TOP features using non-linear registration.**

parcellations	radius	non-BF and LBP-TOP	BF and LBP-TOP	BF and non-BF
AAL	R1	<b>0.0056</b>	<b>0.0102</b>	0.8467
	R2	0.0633	0.1052	
	R3	<b>0.0043</b>	<b>0.0106</b>	
CC200	R1	0.1586	<b>0.0454</b>	<u>0.3619</u>
	R2	<b>0.0093</b>	<b>0.0018</b>	
	R3	<b>0.0013</b>	<b>0.0002</b>	

The  $p$ -values without under lines denote the accuracies of LBP-TOP features are bigger than rs-fMRI data, and the accuracies of rs-fMRI data filtered by a bandpass filter are bigger than rs-fMRI data not filtered. The underlined  $p$ -values show the inverse relationship. Bold  $p$ -values correspond to significant results ( $p$ -value < 0.05). Abbreviations are as follows: **AAL**, automated anatomical labeling template; **CC200**, spatially constrained parcellation based on rs-fMRI; **BF**, rs-fMRI data filtered by a bandpass filter (0.009Hz ~ 0.08Hz); and **non-BF**, rs-fMRI data not filtered by a bandpass filter.



### 5.3.2 Permutation test of basic models

The results of permutation test in Table 5-10 shows each approach can learn the class structure in the data. Classifiers based on LBP-TOP features show more significant than approaches based on rs-fMRI data.

**Table 5-10 Permutation test of some results in Table 5-4**

(a, b) Permutation test of classifiers based on LBP-TOP features

classifier	parcellations	radius	DOF9	DOF12	ART
1NN	brain mask	R1	<b>0.01</b>	<b>0.01</b>	<b>0.01</b>
		R2	<b>0.01</b>	<b>0.01</b>	<b>0.01</b>
		R3	<b>0.01</b>	<b>0.01</b>	<b>0.01</b>
linear-SVM	brain mask	R1	<b>0.01</b>	<b>0.01</b>	<b>0.01</b>
		R2	<b>0.01</b>	<b>0.01</b>	<b>0.01</b>
		R3	<b>0.01</b>	<b>0.01</b>	<b>0.01</b>

(c, d) Permutation test of classifiers based on simple rs-fMRI features

classifier	parcellations	non-BF	BF
1NN	AAL	<b>0.03</b>	<b>0.01</b>
	CC200	<b>0.03</b>	<b>0.04</b>
linea- SVM	AAL	<b>0.03</b>	<b>0.01</b>
	CC200	<b>0.01</b>	<b>0.01</b>

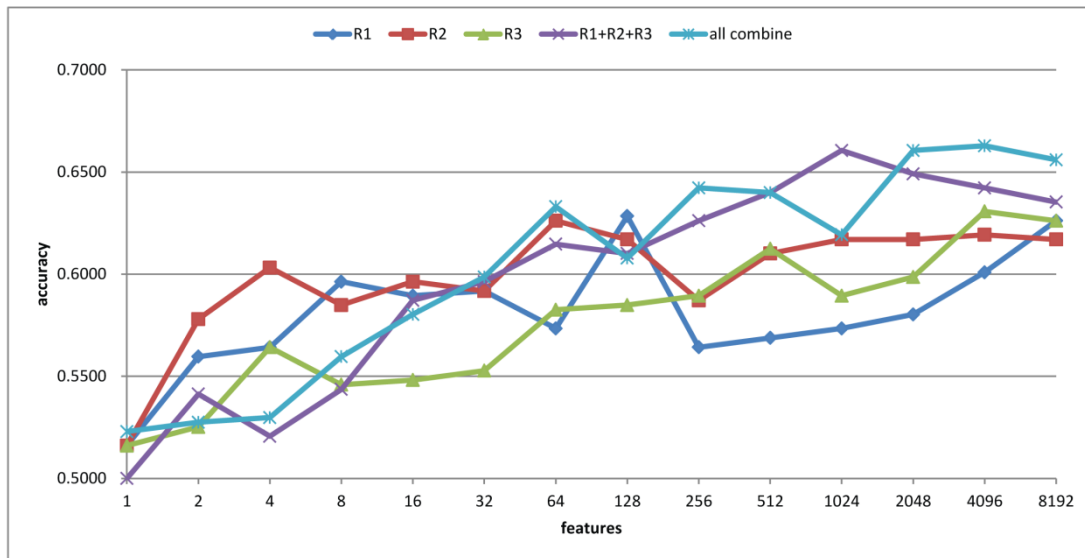
The  $p$ -values are calculated over 100 randomized sets of each test. The error of each set was estimated by same 10-fold cross-validation of data. Bold  $p$ -values correspond to significant results ( $p$ -value < 0.05).



### 5.3.3 *Feature selection*

The feature selection results with the ART non-linear registration methods are shown in Table 5-11. When introducing spatial context information (the AAL and CC200 parcellations), only a few features are needed to build a sufficiently accurate classification model. In most cases, using the same number of features but combining features from all the radii of LBP-TOP (R1+R2+R3) improves the accuracy of the resulting model. After combining all features based on different parcellations and various radii, we achieved a model with greater accuracy compared with the AAL or CC200 parcellations alone. However, the accuracy of the combined model did not surpass that of the model based on the histogram of the whole-brain region. Figure 5-3 shows the test results from using different feature groups based on AAL parcellation.





**Figure 5-3 Feature selection results of ADHD-TDC classification accuracy based on AAL parcellation and the ART non-linear registration method.**

R1, R2, and R3 denote the LBP-TOP radii in mm. R1+R2+R3 denotes the combination all features from R1, R2, and R3. All combine refers to the combination of all features from different parcellations (i.e., brain mask, AAL, and CC200) and various radii.

### 5.3.4 Resolutions of brain images

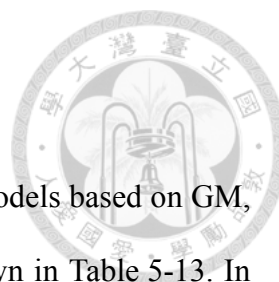
Table 5-12 shows the accuracy of models based on various brain image resolutions. Models utilizing higher resolutions usually had better accuracy. However, models based on the CC200 parcellation had greater accuracy when using 3x3x3 mm resolution. Nevertheless, higher resolution data generally provided more information for the discrimination of ADHD from TDC subjects.



**Table 5-12 Feature selection results of ADHD-TDC classification accuracy of different resolution of source images.**

voxel size	parcellations	features														
		1	2	4	8	16	32	64	128	256	512	1024	2048	4096	8192	
1x1x1 mm	brain mask	0.5183	0.5734	0.6078	0.5917	0.6170	0.6307	0.6147	0.6330	0.6514	<u>0.6583</u>	0.6491	<sup>(531 features)</sup>	0.6422	0.6353	
	AAL	0.5000	0.5413	0.5206	0.5436	0.5872	0.5963	0.6147	0.6101	0.6261	0.6399	<u>0.6606</u>	0.6491	0.6422	0.6353	
	CC200	0.5161	0.5298	0.5528	0.5803	0.6124	0.6101	0.6009	0.5711	0.6147	<u>0.6445</u>	0.6147	0.6353	0.6376	0.6330	
	all combine	0.5229	0.5275	0.5298	0.5596	0.5803	0.5986	0.6330	0.6078	0.6422	0.6399	0.6193	0.6606	<b>0.6628</b>	0.6560	
3x3x3 mm	brain mask	0.4839	0.5528	0.5986	0.5803	0.6032	0.6261	0.6353	0.5940	0.6032	0.6124	0.5986	<sup>(531 features)</sup>	0.6078	0.5963	
	AAL	0.4794	0.4931	0.5000	0.4908	0.4954	0.5252	0.5298	0.5321	0.5505	0.5780	0.6170	0.5780	0.6078	0.5963	
	CC200	0.5092	0.5161	0.4839	0.4679	0.5046	0.5275	0.5183	0.5321	0.6078	0.6284	0.6651	<b>0.6995</b>	0.6972	0.6812	
	all combine	0.4931	0.4794	0.5206	0.5183	0.5046	0.5275	0.5367	0.5573	0.5619	0.6032	0.6422	<u>0.6743</u>	0.6651	0.6583	
6x6x6 mm	brain mask	0.5000	0.5275	0.5505	0.5734	0.5550	0.5619	0.5803	0.5528	0.5436	<u>0.5367</u>	0.5367	<sup>(531 features)</sup>	0.6583	0.6330	
	AAL	0.4725	0.4725	0.4381	0.5046	0.4794	0.4839	0.5183	0.5367	0.5688	0.6032	0.6468	<b>0.6628</b>	0.6583	0.6330	
	CC200	0.4839	0.4885	0.4908	0.5046	0.5092	0.5367	0.5596	0.5482	0.5803	0.5826	0.6055	0.5803	0.5986	<u>0.6193</u>	
	all combine	0.4862	0.4541	0.4771	0.4862	0.4633	0.5206	0.5138	0.5046	0.5390	0.5780	0.5963	0.5940	0.6353	<u>0.6491</u>	




All parcellation (i.e., brain mask, AAL, CC200) results were built by combining LBP-TOP features with R1+R2+R3 and the ART non-linear registration method. All combine denotes the combination of all features from different parcellations (i.e., brain mask, AAL, and CC200) and various radii. The highest accuracy for each resolution is denoted by the bold number. The highest accuracy for each row is underlined. Other abbreviations are as follows: AAL, automated anatomical labeling template; and CC200, spatially constrained parcellation based on rs-fMRI.



### 5.3.5 Tissue types

To determine the most discriminative tissue type within the brain, models based on GM, WM, and CSF probability maps were tested. These results are shown in Table 5-13. In most cases, the structural differences found in the GM data provided the highest discriminative power for separating ADHD from TDC subjects. The McNemar’s test between different tissue types do not show significant difference while using whole brain and AAL parcellations, but show significant difference in some cases using CC200 parcellations (Table 5-14).

**Table 5-13 The ADHD-TDC classification accuracy of models based on the probability map of different brain tissues using the ART non-linear registration method.**

parcellations	radius	features	CSF	WM	GM	whole brain
	R1	177	<b><u>0.6583</u></b>	0.6009	0.6330	0.6514
	R2	177	<u>0.6399</u>	0.6261	0.6330	0.6651
	R3	177	<u>0.6376</u>	0.5986	0.5963	0.6583
	R1	20532	<u>0.6307</u>	0.6147	<u>0.6307</u>	0.6537
	R2	20532	0.6307	<u>0.6353</u>	<u>0.6353</u>	0.6239
	R3	20532	0.6445	0.6284	<b><u>0.6537</u></b>	0.6537
	R1	33630	0.6239	0.6193	<u>0.6560</u>	0.6239
	R2	33630	0.6353	0.6376	<b><u>0.6720</u></b>	0.6537
	R3	33630	0.6399	0.6330	<u>0.6674</u>	0.6697

The highest accuracy obtained for each resolution is noted in bold. The highest accuracy for each row is underlined. Abbreviations are as follows: **R1**, **R2**, and **R3**, the LBP-TOP radii in mm; **AAL**, automated anatomical labeling template; and **CC200**, a spatially constrained parcellation

based on rs-fMRI.



**Table 5-14**  $p$ -values of McNemar’s test of results in Table 5-12

parcellations	radius	GM-CSF	GM-WM
Brain	R1	<u>0.329</u>	0.136
	R2	<u>0.847</u>	0.834
	R3	<u>0.127</u>	<u>1.000</u>
AAL	R1	1.000	0.464
	R2	0.909	1.000
	R3	0.740	0.248
CC200	R1	<b>0.045</b>	0.057
	R2	0.065	0.059
	R3	0.176	0.073

The  $p$ -values without under lines denote the accuracies of GM are bigger than CSF and the accuracies of GM are bigger than WM. The underlined  $p$ -values show the inverse relationship. Bold  $p$ -values correspond to significant results ( $p$ -value < 0.05). **DOF9**, and **DOF12**, linear registration with 9, and 12 degree of freedom, respectively; **ART**, non-linear registration performed by Automated Registration Tool; **AAL**, automated anatomical labeling template; and **CC200**, spatially constrained parcellation based on rs-fMRI.

## 5.4 Discussion

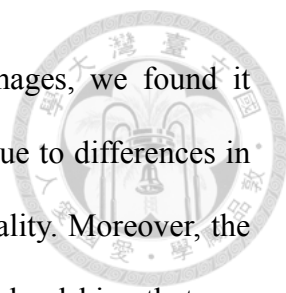
The prevalence of ADHD around the world is highly heterogeneous. Polanczyk et al.

[147] have shown that this variability may be explained primarily by the use of differing ADHD diagnostic criteria and collateral sources of information. Additionally, geographic location also plays a role in the variability of ADHD prevalence around the world [147].

Based on the research of Polanczyk et al., estimations of ADHD prevalence rates using the DSM-III-R or ICD-10 criteria are significantly lower than when using other criteria, such as those of the DSM-IV. Additionally, the use of different collateral sources of information, such as parents, teachers, subjects, the best-estimate procedure, the “and rule (parent and teacher),” or the “or rule (parent or teacher),” can also significantly affect the estimate of ADHD [147].

The ADHD-200 global competition dataset was pooled from research sites all over the world. The organizers of the competition went to great lengths to maintain the consistency of the dataset. Nevertheless, for various historical reasons, including the use of different benchmarks at each site, it is difficult to use the same procedure to diagnose ADHD around the world (Table 5-1). However, the worldwide diagnosis of ADHD reflects an objective reality from which ADHD classification models can be built and evaluated.

While constructing classification models based on machine learning approaches, the inconsistency of diagnostic criteria may introduce so-called class label noise, which may seriously diminish accuracy. Class label noise may be the most important contributor to low accuracy in the ADHD-200 Global Competition.



While constructing our ADHD diagnostic tool based on brain images, we found it difficult to compare the rs-fMRI data from different research sites due to differences in image resolution, slice thickness, time points utilized and image quality. Moreover, the complex preprocessing steps of fMRI data analysis also introduce hardships that can affect the results. Finding the optimal preprocessing strategy to provide the most useful information for building a classifier is a time-consuming process. Therefore, we chose anatomical data rather than rs-fMRI data to mine useful information from brain morphological changes. The resulting classification model based on morphological changes was found to be competitively accurate in discriminating ADHD from TDC subjects. Our results demonstrate that using features based on LBP-TOP data to train the linear SVM can result in greater discriminative power than using features based on rs-fMRI data. The resulting accuracies based on LBP-TOP features are better than those based on rs-fMRI data (Table 5-4).

#### 5.4.1 *Robust to registration method*

The robustness of the registration methods when using LBP-TOP features with ADHD data is notable. Although the model based on the ART non-linear registration method proved to be the most accurate, the models based on linear registrations (FLIRT with 9-DOF and 12-DOF) also performed well in our tests (Table 5-4). This finding demonstrates the stability of the LBP-TOP to registration methods. Due to the large interindividual variability of the human brain, the registration step of MRI brain data analysis is both critical and challenging [61]. Aside from the linear registration method, more than a dozen non-linear registration methods have been developed in recent years,



but a perfect registration method does not yet exist [147].



However, after performing a perfect registration, no structural differences should exist between subjects. Therefore, a good index for morphological changes should not be based on perfect non-linear registration methods. This property of LBP-TOP might provide a simple and efficient way to compare brain morphology with linearly registered brains.

#### 5.4.2 *Global effects of ADHD?*

To introduce different spatial context information, we utilized several parcellation strategies in this study. Unexpectedly, the models using only the distribution of whole-brain features usually demonstrated the highest accuracy in our tests (Table 5-4, Table 5-9, and Table 5-10). Adding parcellation information did not improve the resulting models.

Our results imply that morphological changes in the ADHD brain may affect the whole-brain texture distribution. Further research should be performed to confirm these findings. Theoretically, introducing spatial context information can provide higher accuracy if there are significant structural brain changes in several brain regions. Published structural imaging studies, summarized in two meta-analyses [158, 159], have failed to find robust brain changes between ADHD and control subjects. Meta-analyses can help in identifying brain regions that may be the most abnormal in ADHD subjects. However, it is difficult to build a robust discriminative model of ADHD based only on such selected regions.



### 5.4.3 *Combining models using feature selection*

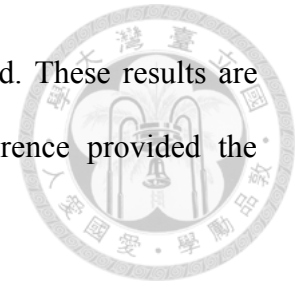
Consider the results of 1NN and linear-SVM in Table 5-4. 1NN uses features as they have same weights, whereas the linear-SVM assigns various weights to them. The results might imply that, with linear registration, use all features with same weight (1NN) cannot provide good results. However, we can make some features more important to make a better classifier (linear-SVM). Only few features might be needed to build a sufficient good classifier in this problem.

To find the most important features and to improve the robustness and efficiency of our model, we used linear-SVM to rank the overall extracted features, and we made an effort to choose the most important features from which to build a better classification model. Moreover, using the feature selection method, we combined models from different point of view to construct a more general model. The results of our tests show that it is useful to combine features to build better models (Table 5-11 and Table 5-12). Moreover, we only need few features to build sufficient good classifiers (Table 5-11). To build a simpler and more robust model, we combined different LBP-TOP features to provide better accuracy. However, when dealing with too many features, the over-fitting effect came into play due to the insufficient number of subjects in this study (436 subjects). In most cases, greater accuracy was not gained by combining more than 4096 features.

### 5.4.4 *Most discriminative tissue*

To determine the most useful brain tissue for discriminating ADHD from TDC subjects,

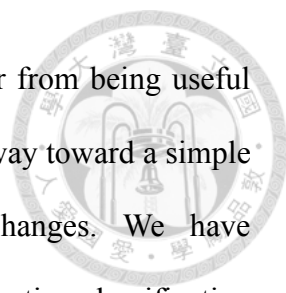
models based on GM, WM, and CSF probability maps were tested. These results are shown in Table 5-13. In most cases, GM-based structural difference provided the greatest discriminative power.



LBP-TOP extracted morphological data based on the distribution of various curvatures, edges, dots, corners, and the content size of the specific region (Figure 5-2). Most of this information may come from the complex patterns of cortical folding, which essentially dominates GM morphology. Therefore, we suggest that the primary morphological information utilized by our model may come from gyrification patterns. Wolosin et al. have previously shown different folding indices for ADHD compared with control subjects [160].

## **5.5 Conclusion**

In this study, we approached the ADHD classification problem by working to find a simple method that could provide sufficient discriminative power. We determined that information derived from texture analysis of brain morphology could be used to distinguish ADHD from TDC subjects. An approach based on structural images is simpler than one based on functional data, and the data are easier to obtain making such an approach potentially more useful in the clinical environment. Our results demonstrate that structural brain data may be another treasure-trove in the ADHD-200 global competition dataset.



Although the accuracy of the models presented in this study are far from being useful clinically, texture difference-based feature extraction may point the way toward a simple and efficient method for determining morphological brain changes. We have demonstrated that LBP-TOP is a good candidate to build a discriminative classification model based on structural brain changes.

## 5.6 Supplements

**Supp. Table 5-1 List of all subjects.**

series	dataset	subject	age	dx	series	dataset	subject	age	dx	series	dataset	subject	age	dx	series	dataset	subject	age	dx
1	KKI	1019436	12.98	3	110	NYU	10069	16.93	0	219	OHSU	2232413	9.17	0	328	Peking_2	3856956	13.75	3
2	KKI	1043241	9.12	0	111	NYU	10070	10.52	0	220	OHSU	2288903	7.42	1	329	Peking_2	3910672	10.08	1
3	KKI	1535233	9.64	0	112	NYU	10071	14.09	3	221	OHSU	2292940	8	1	330	Peking_2	3993793	10.33	0
4	KKI	1577042	9.06	1	113	NYU	10072	7.25	1	222	OHSU	2409220	7.67	0	331	Peking_2	3994098	12.75	0
5	KKI	1594156	12.87	0	114	NYU	10073	14.46	1	223	OHSU	2415970	8.58	3	332	Peking_2	4053388	11.08	1
6	KKI	1686265	8.02	0	115	NYU	10074	8.06	1	224	OHSU	2455205	7.5	1	333	Peking_2	4055710	13.33	3
7	KKI	1779922	10.84	0	116	NYU	10075	11.09	3	225	OHSU	2535204	8.42	1	334	Peking_2	4073815	9.83	3
8	KKI	1842819	10.09	0	117	NYU	10078	8.19	1	226	OHSU	2559559	7.92	3	335	Peking_2	4075719	13.08	1
9	KKI	1988015	11.17	0	118	NYU	10079	15.27	0	227	OHSU	2561174	8.67	1	336	Peking_2	4221029	9.67	1
10	KKI	1996183	11.04	3	119	NYU	10082	11.4	0	228	OHSU	2571197	7.67	1	337	Peking_2	4225073	10.33	3
11	KKI	2014113	10.35	1	120	NYU	10086	15.83	1	229	OHSU	2620872	8.92	1	338	Peking_2	4265987	10.08	0
12	KKI	2018106	11.66	0	121	NYU	10087	17.61	3	230	OHSU	2920716	7.5	0	339	Peking_2	5993008	12.25	3
13	KKI	2104012	9.62	0	122	NYU	10088	14.66	0	231	OHSU	2947936	8.5	0	340	Peking_2	6500128	11.33	3
14	KKI	2299519	11.89	1	123	NYU	10090	16.32	1	232	OHSU	3051944	11.75	0	341	Peking_2	7011503	13.33	0
15	KKI	2360428	8.56	0	124	NYU	10091	8.93	1	233	OHSU	3052540	9.17	1	342	Peking_2	7253183	13.17	3
16	KKI	2554127	9.02	0	125	NYU	10092	17.3	0	234	OHSU	3162671	8.5	0	343	Peking_2	7407032	13.42	0
17	KKI	2558999	8.87	0	126	NYU	10095	14.39	1	235	OHSU	3244985	7.17	0	344	Peking_2	7689953	13.5	3
18	KKI	2572285	8.33	0	127	NYU	10096	14.94	3	236	OHSU	3302025	9	0	345	Peking_2	8278680	13.25	3
19	KKI	2601925	9.58	0	128	NYU	10100	16.88	0	237	OHSU	3466651	8.08	1	346	Peking_2	9002207	11.17	1
20	KKI	2618929	10.09	0	129	NYU	10101	14.2	1	238	OHSU	3470141	8.5	1	347	Peking_2	9578631	14.33	0
21	KKI	2640795	12.47	0	130	NYU	10102	16.55	0	239	OHSU	3652932	7.67	3	348	Peking_2	9640133	13.75	0
22	KKI	2703289	11.1	0	131	NYU	10103	9.59	1	240	OHSU	3677724	8.75	3	349	Peking_3	1050345	12.67	0
23	KKI	2740232	11.09	0	132	NYU	10104	13.35	3	241	OHSU	3869075	8.67	0	350	Peking_3	1132854	13.92	0
24	KKI	2917777	12.66	0	133	NYU	10106	11.82	3	242	OHSU	4016887	9.08	1	351	Peking_3	1356553	11.75	1
25	KKI	2930625	9.97	0	134	NYU	10107	17.37	3	243	OHSU	5302451	11.33	3	352	Peking_3	1399863	12.83	1
26	KKI	3103809	9.3	0	135	NYU	10111	7.74	0	244	OHSU	6592761	7.58	0	353	Peking_3	1404738	12.58	0
27	KKI	3119327	10.15	3	136	NYU	10112	14.79	0	245	OHSU	6953386	11.83	1	354	Peking_3	1411536	13.17	0
28	KKI	3154996	11.65	3	137	NYU	10114	16.73	0	246	OHSU	7333005	9	1	355	Peking_3	1662160	12.58	0
29	KKI	3160561	11.95	1	138	NYU	10115	15.32	3	247	OHSU	8218392	8.33	0	356	Peking_3	1771270	14.42	3
30	KKI	3611827	9.08	1	139	NYU	10116	16.77	3	248	OHSU	9499804	8.17	0	357	Peking_3	1794770	11.75	0
31	KKI	3699991	12.77	0	140	NYU	10117	9.35	0	249	Peking_1	1056121	13.92	0	358	Peking_3	1843546	14.08	3
32	KKI	3884955	11.84	0	141	NYU	10118	13.33	1	250	Peking_1	1133221	12.33	1	359	Peking_3	2107404	13.08	0
33	KKI	3902469	10.64	0	142	NYU	10119	7.77	3	251	Peking_1	1186237	13.92	3	360	Peking_3	2208591	12.5	0
34	KKI	3917422	8.19	1	143	NYU	10122	16.28	0	252	Peking_1	1240299	9.5	1	361	Peking_3	2228148	13.75	1
35	KKI	3972956	10.12	0	144	NYU	10123	16.13	0	253	Peking_1	1282248	10.58	3	362	Peking_3	2268253	14.83	0
36	KKI	4104523	12.43	0	145	NYU	10125	7.19	0	254	Peking_1	1408093	11.92	0	363	Peking_3	2276801	13	3
37	KKI	5216908	10.82	0	146	NYU	10126	9.7	1	255	Peking_1	1879542	13.58	0	364	Peking_3	2493190	13.33	0
38	KKI	6346605	10.69	0	147	NYU	10128	9.53	0	256	Peking_1	1947991	9.42	3	365	Peking_3	2524687	14.33	3
39	KKI	7129258	9.73	0	148	NYU	1000804	7.29	0	257	Peking_1	2174595	9.83	1	366	Peking_3	2780647	13.58	1
40	KKI	7415617	10.9	0	149	NYU	1023964	8.29	3	258	Peking_1	2196753	10.33	3	367	Peking_3	2907951	16	3
41	KKI	8083695	8.62	0	150	NYU	1057962	8.78	1	259	Peking_1	2266806	11.25	0	368	Peking_3	2940712	13.08	3
42	KKI	8263351	8.34	0	151	NYU	1187766	12.79	1	260	Peking_1	2367157	12.08	1	369	Peking_3	2984158	14.17	3
43	KKI	8337695	8.8	1	152	NYU	1283494	8.61	1	261	Peking_1	2535087	13.33	0	370	Peking_3	3224401	12.33	0
44	KKI	8658218	9.75	0	153	NYU	1320247	8.59	0	262	Peking_1	2833684	10.42	0	371	Peking_3	3277313	12.92	0
45	KKI	9922944	10.3	0	154	NYU	1359325	10.76	0	263	Peking_1	2910270	13.83	3	372	Peking_3	3291029	14.83	3
46	NeuroIMAGE	1125505	19.3	0	155	NYU	1435954	11.92	0	264	Peking_1	3004580	14	0	373	Peking_3	3385520	13.5	0
47	NeuroIMAGE	1312097	15.41	0	156	NYU	1471736	13.32	3	265	Peking_1	3086074	11.5	0	374	Peking_3	3473830	12.67	0
48	NeuroIMAGE	1585708	15.96	1	157	NYU	1497055	8.56	1	266	Peking_1	3212536	12.17	0	375	Peking_3	3624598	14.58	0
49	NeuroIMAGE	2029723	16.5	1	158	NYU	1511464	10.12	3	267	Peking_1	3269608	12.08	0	376	Peking_3	3672300	14	3
50	NeuroIMAGE	2352986	20.47	0	159	NYU	1517240	10.06	1	268	Peking_1	3390312	10.33	3	377	Peking_3	3712305	11	3
51	NeuroIMAGE	2671604	17.46	0	160	NYU	1737393	11.24	0	269	Peking_1	3593327	10	0	378	Peking_3	3803759	13.17	3
52	NeuroIMAGE	2961243	16.51	1	161	NYU	1780174	11.18	1	270	Peking_1	3707771	10.42	0	379	Peking_3	3870624	11	1
53	NeuroIMAGE	3007585	19.05	0	162	NYU	1884448	8.91	0	271	Peking_1	3889095	11.92	0	380	Peking_3	3930512	12.58	0
54	NeuroIMAGE	3048588	13.92	1	163	NYU	1918630	7.55	1	272	Peking_1	3976121	11	1	381	Peking_3	4006710	14	3

55	NeuroIMAGE 3082137	17.93	1	164	NYU 1992284	7.26	1	273	Peking_1 3983607	9.67	1	382	Peking_3 4048810	14.67	0
56	NeuroIMAGE 3108222	13.33	1	165	NYU 1995121	11.43	0	274	Peking_1 4028266	9.25	3	383	Peking_3 4136226	11.25	0
57	NeuroIMAGE 3304956	15.78	0	166	NYU 2030383	13.56	1	275	Peking_1 4095748	10.5	1	384	Peking_3 4241194	11.67	1
58	NeuroIMAGE 3322144	17.5	1	167	NYU 2054438	8.11	1	276	Peking_1 4334113	14.58	3	385	Peking_3 5575344	13.75	0
59	NeuroIMAGE 3449233	20.15	1	168	NYU 2107638	10.41	1	277	Peking_1 4921428	9.58	0	386	Peking_3 5669389	13.75	0
60	NeuroIMAGE 3515506	18.8	3	169	NYU 2136051	9.23	0	278	Peking_1 7093319	11.42	0	387	Peking_3 6383713	11.75	1
61	NeuroIMAGE 3566449	17.78	1	170	NYU 2230510	9.17	1	279	Peking_1 7390867	17.33	3	388	Peking_3 6477085	13	0
62	NeuroIMAGE 3808273	14.25	1	171	NYU 2260910	12.07	1	280	Peking_1 9210521	9.25	3	389	Peking_3 7994085	14.92	0
63	NeuroIMAGE 3858891	18.32	0	172	NYU 2297413	11.06	3	281	Peking_1 9221927	11.83	0	390	Peking_3 8191384	13.42	0
64	NeuroIMAGE 3888614	16.8	1	173	NYU 2306976	11.53	1	282	Peking_1 9887336	11.33	0	391	Pittsburgh 16001	10.11	0
65	NeuroIMAGE 3941358	18.2	1	174	NYU 2497695	11.61	1	283	Peking_2 1050975	13.58	0	392	Pittsburgh 16003	10.46	0
66	NeuroIMAGE 3980079	19.15	1	175	NYU 2570769	13.49	1	284	Peking_2 1068505	10.25	0	393	Pittsburgh 16004	10.49	0
67	NeuroIMAGE 4134561	15.55	0	176	NYU 2682736	10.78	1	285	Peking_2 1093743	11.92	0	394	Pittsburgh 16006	10.76	0
68	NeuroIMAGE 4239636	20.54	1	177	NYU 2735617	11.42	0	286	Peking_2 1094669	12.42	1	395	Pittsburgh 16008	11.03	0
69	NeuroIMAGE 4285031	12.84	0	178	NYU 2741068	8.53	1	287	Peking_2 1117299	13.5	0	396	Pittsburgh 16009	11.12	0
70	NeuroIMAGE 6115230	19.39	0	179	NYU 2821683	10.69	3	288	Peking_2 1159908	15.08	1	397	Pittsburgh 16010	11.24	0
71	NeuroIMAGE 8409791	17.59	0	180	NYU 2950672	11.28	1	289	Peking_2 1341865	9.25	1	398	Pittsburgh 16012	11.41	0
72	NYU 10009	12.1	0	181	NYU 2991307	8.3	0	290	Peking_2 1494102	10.33	0	399	Pittsburgh 16013	11.53	0
73	NYU 10010	11.32	0	182	NYU 2996531	12.62	3	291	Peking_2 1562298	9.58	0	400	Pittsburgh 16014	11.69	0
74	NYU 10015	11.21	3	183	NYU 3163200	11.91	0	292	Peking_2 1628610	14.92	1	401	Pittsburgh 16015	11.7	0
75	NYU 10017	12.38	1	184	NYU 3174224	8.49	1	293	Peking_2 1643780	10	1	402	Pittsburgh 16016	11.71	0
76	NYU 10018	11.54	1	185	NYU 3349423	12.59	0	294	Peking_2 1809715	13.5	3	403	Pittsburgh 16021	12.41	0
77	NYU 10019	14.81	1	186	NYU 3433846	10.36	1	295	Peking_2 1860323	8.75	0	404	Pittsburgh 16025	12.71	0
78	NYU 10022	12.12	1	187	NYU 3441455	9.26	3	296	Peking_2 1875013	9.83	0	405	Pittsburgh 16026	12.88	0
79	NYU 10023	13.42	0	188	NYU 3619797	7.49	1	297	Peking_2 1916266	13.17	0	406	Pittsburgh 16027	12.91	0
80	NYU 10024	8.26	0	189	NYU 3653737	9.38	1	298	Peking_2 2031422	10.92	3	407	Pittsburgh 16028	13.03	0
81	NYU 10025	8.75	1	190	NYU 3679455	11.97	1	299	Peking_2 2033178	11.33	0	408	Pittsburgh 16029	13.11	0
82	NYU 10026	11.73	1	191	NYU 3845761	8.19	0	300	Peking_2 2140063	9.42	0	409	Pittsburgh 16030	13.44	0
83	NYU 10028	9.42	3	192	NYU 4060823	11.8	3	301	Peking_2 2141250	12.83	3	410	Pittsburgh 16031	13.53	0
84	NYU 10030	12.41	3	193	NYU 4079254	10.19	0	302	Peking_2 2207418	13.17	1	411	Pittsburgh 16033	13.68	0
85	NYU 10032	11.1	1	194	NYU 4095229	9.32	1	303	Peking_2 2296326	12.17	0	412	Pittsburgh 16035	13.94	0
86	NYU 10033	10.97	3	195	NYU 4154672	9.23	1	304	Peking_2 2310449	10.33	0	413	Pittsburgh 16036	13.97	0
87	NYU 10035	8.95	3	196	NYU 4164316	10.69	0	305	Peking_2 2377207	13.83	0	414	Pittsburgh 16037	14.15	0
88	NYU 10037	10.9	1	197	NYU 4187857	13.28	1	306	Peking_2 2498847	9.75	0	415	Pittsburgh 16038	14.24	0
89	NYU 10039	13.28	0	198	NYU 5971050	7.24	1	307	Peking_2 2529026	13.33	3	416	Pittsburgh 16039	14.57	0
90	NYU 10040	9.94	1	199	NYU 8692452	12.25	0	308	Peking_2 2559537	11.17	0	417	Pittsburgh 16044	14.85	0
91	NYU 10042	10.65	1	200	NYU 8697774	9.92	1	309	Peking_2 2601519	13.83	3	418	Pittsburgh 16048	15.56	0
92	NYU 10045	12.49	0	201	NYU 8834383	8.8	0	310	Peking_2 2659769	14.25	0	419	Pittsburgh 16050	15.78	0
93	NYU 10046	9.75	0	202	NYU 8915162	10.7	1	311	Peking_2 2737106	14.17	3	420	Pittsburgh 16052	15.91	0
94	NYU 10047	12.72	1	203	NYU 9326955	10.92	1	312	Peking_2 2884672	12.83	3	421	Pittsburgh 16053	15.92	0
95	NYU 10048	14.95	1	204	NYU 9750701	10.74	0	313	Peking_2 2919220	13.58	3	422	Pittsburgh 16054	16.05	0
96	NYU 10049	7.35	1	205	NYU 9907452	9.98	3	314	Peking_2 2950754	13.33	1	423	Pittsburgh 16055	16.16	0
97	NYU 10050	17.43	1	206	OHSU 1084283	11	1	315	Peking_2 3157406	14	0	424	Pittsburgh 16060	17.16	0
98	NYU 10051	15.4	3	207	OHSU 1108916	8.5	1	316	Peking_2 3194757	12.92	3	425	Pittsburgh 16061	17.23	0
99	NYU 10052	16.31	0	208	OHSU 1206380	9.17	3	317	Peking_2 3205761	14.58	1	426	Pittsburgh 16063	17.46	0
100	NYU 10054	17.83	0	209	OHSU 1386056	8	0	318	Peking_2 3248920	12.17	0	427	Pittsburgh 16064	17.49	0
101	NYU 10056	15.63	3	210	OHSU 1411223	9.17	1	319	Peking_2 3308331	9.08	0	428	Pittsburgh 16072	18.43	0
102	NYU 10057	17.7	0	211	OHSU 1418396	8.67	0	320	Peking_2 3446674	14.58	3	429	Pittsburgh 16073	18.48	0
103	NYU 10058	14.2	0	212	OHSU 1421489	8.75	0	321	Peking_2 3494778	9.17	0	430	Pittsburgh 16078	18.72	0
104	NYU 10060	8.75	1	213	OHSU 1481430	11.92	0	322	Peking_2 3561920	13.75	1	431	Pittsburgh 16079	18.74	0
105	NYU 10061	11.41	3	214	OHSU 1548937	11.17	0	323	Peking_2 3562883	11.25	0	432	Pittsburgh 16081	18.81	0
106	NYU 10062	16.14	3	215	OHSU 1679142	10.33	0	324	Peking_2 3610134	10.83	0	433	Pittsburgh 16083	18.86	0
107	NYU 10064	15.9	1	216	OHSU 1743472	8.75	1	325	Peking_2 3655623	13.83	0	434	Pittsburgh 16084	18.9	0
108	NYU 10067	13.1	3	217	OHSU 2054310	8.42	1	326	Peking_2 3691107	9.67	1	435	Pittsburgh 16086	18.93	0
109	NYU 10068	13.3	0	218	OHSU 2071989	8.25	1	327	Peking_2 3827352	15.83	3	436	Pittsburgh 16087	18.96	0

Abbreviations are as follows: **dx**, diagnosis; **0**, typically developing children (TDC); **1**, ADHD combined-type; **3**, ADHD inattentive-type; **KKI**, the Kennedy Krieger Institute; **NeuroIMAGE**, the NeuroIMAGE sample; **NYU**, the New York University Child Study Center; **OHSU**, Oregon Health and Science University; **Peking\_1**, **Peking\_2**, and **Peking\_3**, Peking University; and **Pittsburgh**, the University of Pittsburgh.



## Chapter 6

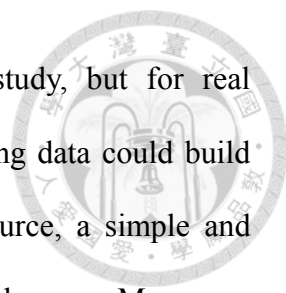
### Discussion, Conclusion, and Future Works

#### 6.1 Discussion

##### 6.1.1 *Structural MRI*

In this study, we used LBP-TOP as a tool to extract information from T1WI and DTI (Chapter 4 and Chapter 5). Results showed LBP can perform on both types of brain MR images and LBP histogram can be used to construct either classification (Chapter 5) or regression models (Chapter 4). There are two important properties of LBP-TOP in structural MRI applications.

First, simple and efficient, computing the LBP histograms of specific brain region is very simple and efficient. The algorithm is nothing more than comparing the scale value of each voxels and its neighbors. Simple and efficient is very important in big data

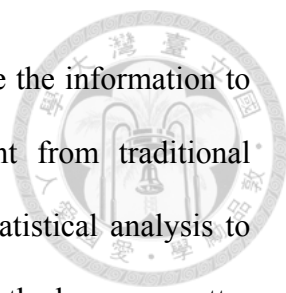


approaches. We only considered hundreds of subjects in this study, but for real applications, using thousands or even millions of subjects as training data could build more accurate results. Therefore, with limited computational resource, a simple and effective information extraction method is more useful than a complex one. Moreover, as wrote by Google's artificial-intelligence guru, Peter Norvig, “*Simple models and a lot of data trump more elaborate models based on less data.*” [2], we can expect the results accuracy of LBP is better than other complex methods.

Second, robust to registration method, as we showed in 5.4.1 and Table 5-4, LBP-TOP is robust to the registration method used for normalizing original brain MR images to standard templates. Because LBP only considers the relative value of the eight neighbors and central voxel, this pattern is naturally robust to slight changes of rotation or scaling. More interesting, based on this property of LBP, we tried to apply LBP-TOP on brain MR images without registration, that is, the unnormalized brain images. Results in Chapter 4 show that LBP-TOP can effectively extract useful information from unnormalized brain images and provide excellent performance (Table 4-9). This property is very important for LBP in the brain MR image applications because the steps of registration are still critical and almost impossible to verify [61] [147]. A method without registration is more useful for comparing data across centers all over the world.

### 6.1.2 *Different from Traditional Approaches*

Based on the proposed methods, binary pattern extract structural information from



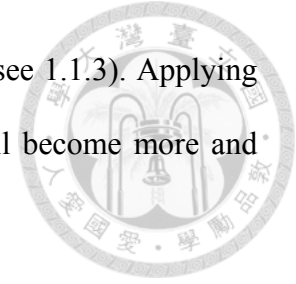
different brain area and then the machine learning algorithm combine the information to construct a discriminative model. The process is quite different from traditional approaches, such as VBM. While performing VBM, we use the statistical analysis to construct statistical inference based on normalized and smoothed gray matter concentration extracted from brain images. The inference based on stochastic data model shows which part of brain area is more "important" in mental disorders. Those studies are good for find starting points of further research. It is hard to build discriminative rules in clinical use based on the results of those studies.

The proposed methods used machine learning algorithms to combine related factors and construct discriminative models to predict or estimate. There is no underlying data model behind this approach. The inference is based on the information provided from the training data, so called data driven approaches. Moreover, the results are ready to use in clinical. Therefore, proposed methods are more useful to build automatic computer-aided diagnosis tools in clinical environment.

Over the past decade, most researchers can only collect few subject data (dozens or hundreds of subjects). Traditional approaches are suitable for few sample size. However, the results from small subjects are usually hard to be constant across studies when apply on multifactorial and heterogeneous disorders, such as schizophrenia [161, 162]. In big data era, it is possible to get much more data than traditional approaches. Therefore, use machine learning and simple information extraction methods to build ready to used models is more reasonable and useful. Currently, it is technologically possible to pool all digitalized brain MR data stored in hospitals all over the world. Actually, many



institutions are trying hard to accomplish this dream in the future (see 1.1.3). Applying machine learning approached on millions of brain MR images will become more and more important in the future.



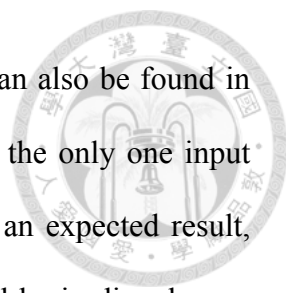
### 6.1.3 *Advantages of Using Binary Patterns and Machine Learning Approaches*

Machine learning algorithms are designed for solve complex problems with multiple contributing factors. Therefore, the introduced method in this study is best suit for analyzing multifactorial and heterogeneous disorders, such as ADHD, schizophrenia, autism, and so on. Moreover, it is also useful in analyzing the continuous process combined with related effects, such as brain maturation and aging.

There are three advantages of using LBP-TOP to extract information from structural brain MR images: First, because of the robustness of LBP-TOP to the registration methods of images and only considering the relative value, LBP-TOP is good for analyzing data pooled from various data centers. Second, as shown in previous chapters, the rank of the features of learned models can be used to valuate further research directions. Third, the proposed methods build a ready to used model as the training result, which could be used as the screening step after each image acquisition in clinical usage.

### 6.1.4 *Knowledge Discovery*

In all three tests of this study, we tried to identify the most important features in each



learned model. Many top ranked brain areas based on the models can also be found in the literature. However, if we only use one of those brain areas as the only one input features, the learned results never provide good performance. It is an expected result, because the target problems analyzed in this study are multifactorial brain disorders or heterogeneous processing.

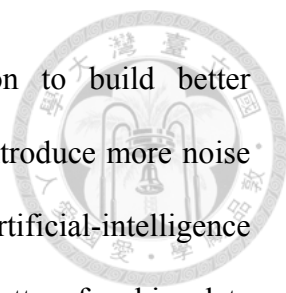
### 6.1.5 *Limitations*

#### *Limitations of Binary Patterns Approaches*

In this study, we used LBP for structural MRI and build several functional connectivity binary patterns for rs-fMRI. Those methods convert the information embedded in the original data to a simple pattern distribution. Although we have shown the usability of this approach, there are several limitations.

First, the pattern distribution is good for machine learning methods to learn a discriminated model but bad for human beings to understand the resulted models. Therefore, we can only find the rank of the discriminated area in the brain based on the learned models. It is hard to figure out why or how the model uses those patterns to build a model.

Second, binary patterns are a simplified form of the information embedded in original data. While building binary patterns, we dropped much information. We only considered the interactions of pixels in three orthogonal planes while constructed the LBP-TOP.



More complex forms might introduce more useful information to build better discriminated models. However, more complex forms might also introduce more noise and provide worse models. Following the words wrote by Google artificial-intelligence guru Peter Norvig [2], binary pattern approaches could be better for big data applications than statistical analysis in small groups.

### *Limitations of Big Data Approaches*

One of the major shifts of mindset in big data era is to use almost all the data rather than using a small number of sample size [22]. Though we used hundreds of subjects to build the discriminated models in this study, it is far from the total number of human beings. Moreover, as we showed in this study, age and sex affect the brain structure. Much more data and more attributes of each subject can help the machine learning algorithms to construct better models. In the current state, the number of the used subjects is the most important limitation of big data approaches. Our studies here only showed the possibility and usability of using binary patterns to build discriminated models. For real clinical usage, much more data from different hospitals and research centers is needed.

## 6.2 Future Works



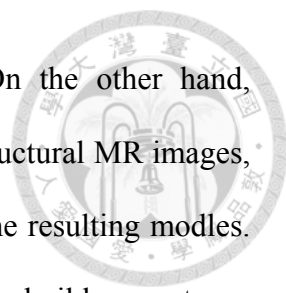
### 6.2.1 *Normal Ranges as Image Biomarkers of Brain Images*

Besides using the binary patterns as the input features for machine learning methods, the quantified and compared distances can be computed based on different histogram [163]. We can use those distances to summarize the information to simpler medical indexes based on brain structural MRI data. Moreover, we can build the “normal range” of structural MRI based on that index. A simpler index can be better understood and used as an image biomarker in clinical environment for quickly screening and be easily combined with the information from other medical examination results.

However, based upon the results in Chapter 4, age does highly affect the brain structure. The index can be more useful and reasonable in same-age groups. Therefore, the Alzheimer's disease neuroimaging initiative (ADNI) database [164] could be a good target to test this idea. We will try to design and evaluate simple medical index based on distances of binary patterns approaches to evaluate Alzheimer's disease and test it using ADNI database.

### 6.2.2 *Combine Information for Multivariate Approaches*

Although only using the brain MRI data can acquire good accuracy in several applications. Brain MR image is a small part of patient information. While learning classification models, the result accuracy would be better if we can combine almost all available patient information, such as the demographic and genetic characteristics,



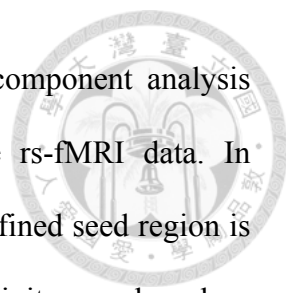
physiological constants, and other medical examination results. On the other hand, combining informations from different brain MR images, such as structural MR images, DTI, and functional MRI, could also improve the performance of the resulting modes. Moreover, hundreds of parents used in this study are not enough to build very strong inferences. The result could be better while learning classification or regression modes from millions of subjects.

### 6.2.3 *Detect ADHD and Schizophrenia Using Functional Connectivity*

#### *Binary Patterns*

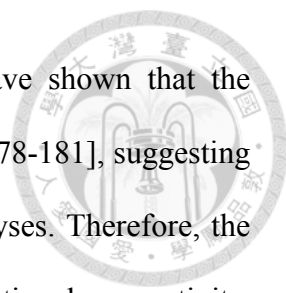
In recent years, resting-state functional magnetic resonance imaging (rs-fMRI) has become a novel technique for studying mental illnesses [12-15]. However, developing a simple and effective method to extract information from rs-fMRI data remains critical. For example, the examination of schizophrenia shows different patterns in the resting-state functional connectivity between patients and healthy controls, but these results were not constant across studies [161, 162]. In addition, researchers have identified several abnormal rs-fMRI patterns in attention deficit-hyperactivity disorder (ADHD) [165], but classifying ADHD using a large pooled rs-fMRI database still requires improvement [36]. The complexity of resting-state functional connectivity increases the difficulty of obtaining consistent results. Therefore, we proposed binary pattern distribution as an approach to improve performance when discriminating individuals with mental disorders from normal control subjects.

Currently, there are two general methods for studying the functional connectivity in

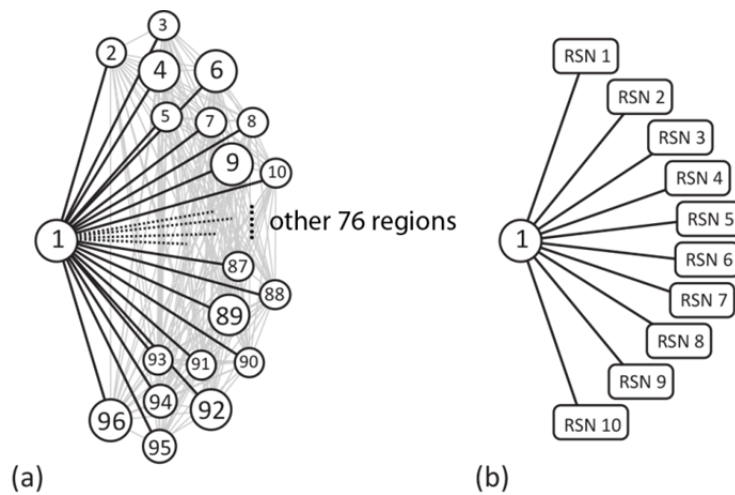


rs-fMRI data, cross-correlation analysis (CCA) and independent component analysis (ICA)[166]. CCA is the most popular method used to analyze rs-fMRI data. In seed-based CCA (SCCA), the average time series in a previously defined seed region is used as a reference. Subsequently, a whole-brain functional connectivity map based on the seed is computed after considering the correlation between each voxel and the reference time series. SCCA is simple method, but using few seeds likely shows only small regions of the whole-brain functional connectivity. Thus, the graph analysis framework [167], such as network CCA (NCCA), was introduced. In NCCA, the whole-brain volume is separated into many regions and the average time-series of each region is calculated. Subsequently, a network is used to describe whole-brain functional connectivity after computing the correlations of each paired regions. Researchers have used these connections as features to obtain classification models with machine learning methods, for determining brain maturation [168], for the classification of autism [169], and for the identification of dementia [170].

In contrast, ICA is a data driven, hypothesis-free method. ICA decomposes the rs-fMRI data into a number of statistically significant spatially independent patterns [171-174]. Independent components (ICs) are considered as resting-state networks (RSNs), which are consistent across health subjects [175]. More importantly, some major ICA patterns are similar to the patterns of activation maps derived from a large database of event-related fMRI studies [1]. Nevertheless, ICA does not offer any information about the intrinsic order of the ICs. However, it might be difficult to determine why and which IC should be chosen as the target of analysis [176, 177].



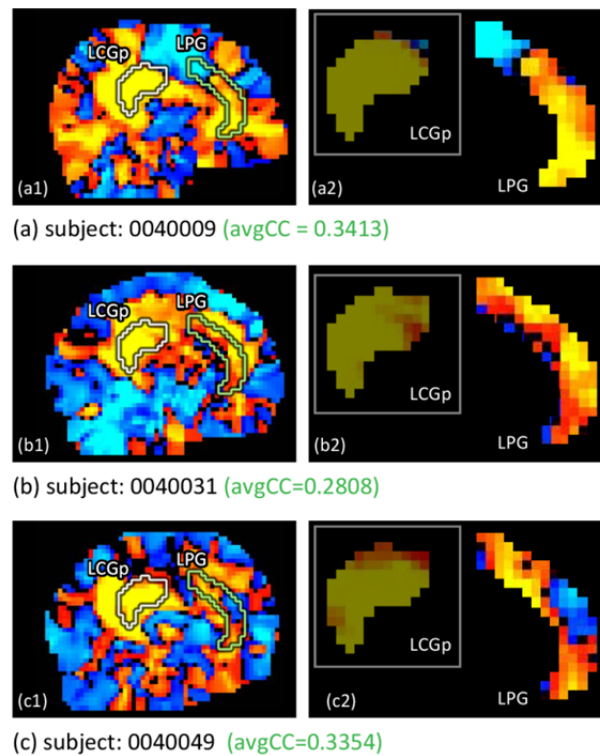
Although CCA and ICA are different approaches, researchers have shown that the results are similar, primarily generating overlapping patterns [166, 178-181], suggesting that the same underlying connectivity structure exists in these analyses. Therefore, the RSNs derived from ICA could be a good reference to analyze functional connectivity. The basic assumption of our strategy is that most RSN patterns are similar across subjects. Therefore, the time variation of each RSN could be a good reference to analyze functional connectivity. Furthermore, the connectivity of each region could be established through the correlations between average time course of region and RSN time series, herein referred to as the resting-state networks referenced analysis (RSNRA)(Figure 6-1).



**Figure 6-1 Simplify connectivity using resting-state networks referenced analysis (RSNRA).**

- (a)** For network cross-correlation analysis with 96 brain regions to construct the connectivity of region 1, all 96 correlations should be considered. **(b)** In RSNRA, only N correlations between the region 1 and N reference resting-state networks (RSNs) should be considered (in this case, N=10).

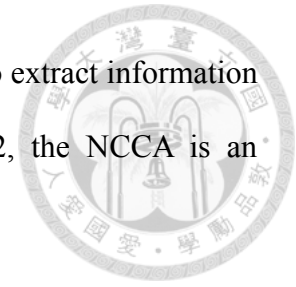




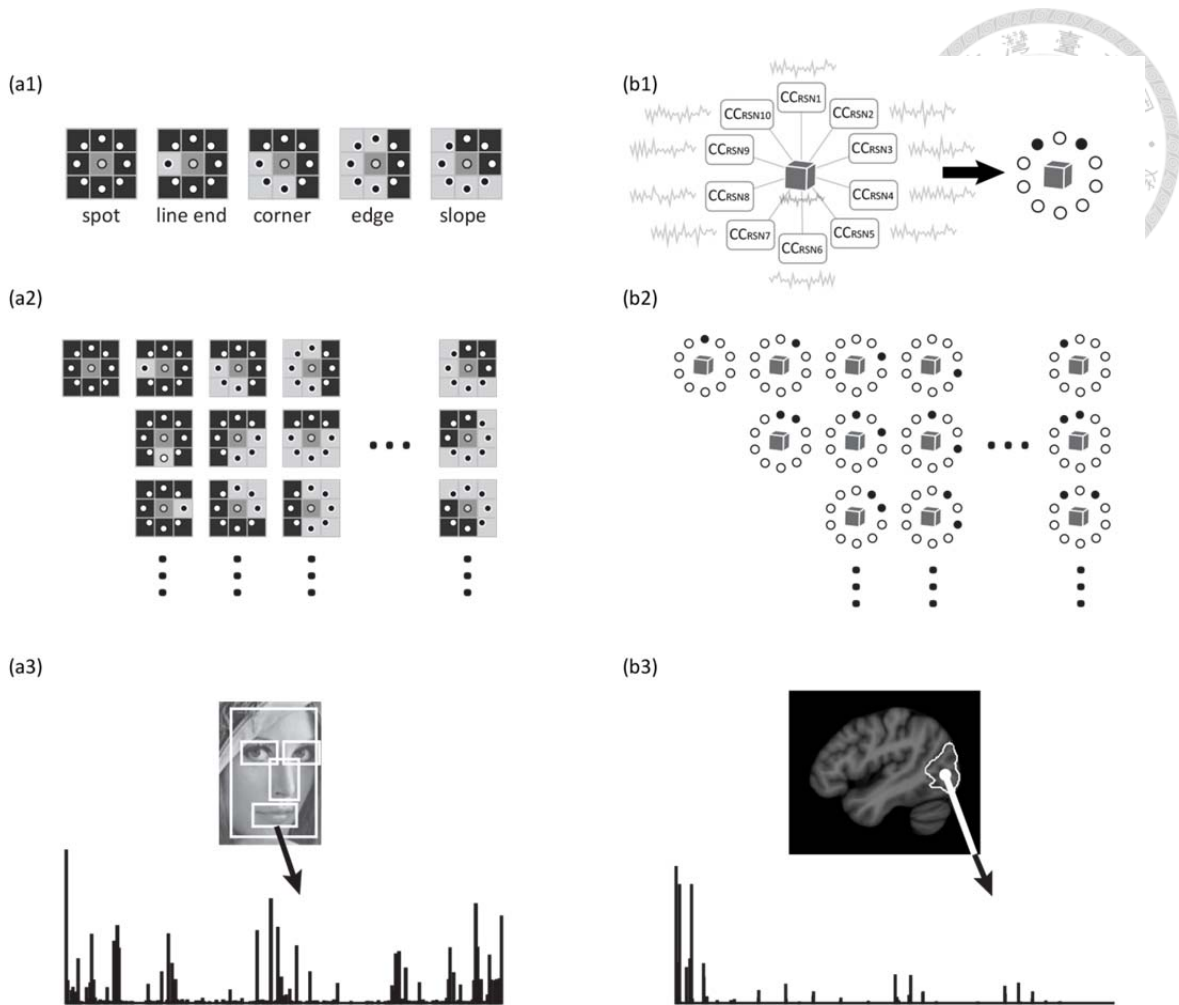
**Figure 6-2 The distribution information not included in the network cross-correlation analysis.**

This figure illustrates the functional connectivity maps of three subjects from the COBRE database. The maps were built using left cingulate gyrus, posterior division (LCGp) as the seed. The subjects have similar average Pearson's correlation coefficients (avgCC) between the average time series of the left paracingulate gyrus (LPG) and average time series of LCGp. Images a1, b1, and c1 show functional connectivity maps of the whole brain that were obtained by calculating the CC of the time series of each brain voxel and the average time series of the seed. Images a2, b2, and c2 show the details of different distribution patterns existing in LPG. These distinct distribution patterns were not considered in the network cross-correlation analysis.

Because of the simple implementation, studies usually use NCCA to extract information from rs-fMRI data [168-170]. However, as shown in Figure 6-2, the NCCA is an average approach without different distribution information.

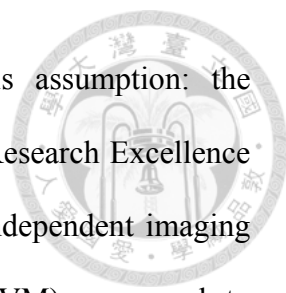


The distribution information is generally used in the computer vision domain to describe objects for determination and classification [62, 73], as we have shown in 2.1. In LBP, the characteristics of each pixel are defined as the types of binary comparison between each pixel and its neighbors [73]. For example, the spots, line ends, and corners are represented as different types of patterns (Figure 6-3 (a1)). Thus, the distribution of these patterns could provide an excellent information source to describe the objects in the images (Figure 6-3 (a3)). Based on this idea, we used RSNRA to describe the connectivity of each brain voxel and extract discriminative information from the distribution of different functional connectivity binary patterns (FCBP) in different brain regions (Figure 6-3 (b1) and (b3)). For this purpose, we introduced and tested the threshold (TFCBP), ordered (OFCBP), and absolute ordered (ABS-OFCBP) FCBP to construct an FCBP histogram of each brain region. The analysis shown in Figure 6-3 compares the difference between LBP and proposed FCBP.



**Figure 6-3 The construction of local binary patterns (LBP) and functional connectivity binary patterns (FCBP).**

**(a1)** In LBP, the characters of each pixel are defined as the types of the binary comparison between each pixel and its neighbors. If the value of the neighbor is larger than the central pixel, the result would be one (filled circle); otherwise, the result is zero (empty circle). Therefore, several types of pattern can be encoded into the binary pattern. **(a2)** All the possible patterns could be used to describe the characters of specific image area. **(a3)** The LBP histogram can then be used as the feature for image classification. **(b1)** In FCBP, the characters of each voxel are defined as the cross correlation between the time series of the voxel and the time courses of the referenced resting-state networks ( $CC_{RSN1} \sim CC_{RSN10}$ ). **(b2)** Based on two binarization strategies we introduced in this study, all possible FCBP could be used to describe the characters of the specific brain region. **(b3)** Therefore, the FCBP histogram can be used as the feature for brain classification.



Two public-access rs-fMRI databases were used to assess this assumption: the schizophrenia database contributed by The Center for Biomedical Research Excellence (COBRE) and ADHD-200 Sample database obtained from eight independent imaging sites. For each database, the linear support vector machine (SVM) was used to determine the discriminative models for classifying patients and control subjects using the extracted features based on different approaches. We examined and compared the performance of the constructed models using the features extracted from rs-fMRI data through NCCA, RSNRA, and FCBP.

### *Materials and Methods*

The parcellations and evaluations methods are the same with 3.1.

#### ***Participants***

Rs-fMRI data from two public-access databases in the International Neuroimaging Data-sharing Initiative (INDI) were used to evaluate the methods proposed in this study [182]. To rule out sexual differences, only male subjects were used. Subjects with translation movements larger than 3 mm or rotations larger than 3 degrees were ruled out. Because cerebellum RSN patterns are used in the study, cerebellum volumes less than 65% in the registered brain data were removed.

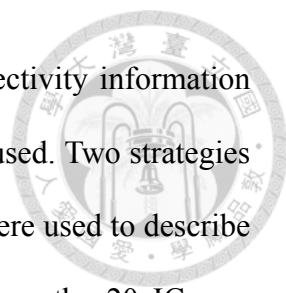
The Schizophrenia database was obtained from The Center for Biomedical Research Excellence (COBRE). COBRE includes 72 patients with schizophrenia and 74 healthy

controls. Rs-fMRI images, anatomical MRI volumes, the detail phenotypic data, and diagnostic information of each subject can be freely download from [http://fcon\\_1000.projects.nitrc.org/indi/retro/COBRE.html](http://fcon_1000.projects.nitrc.org/indi/retro/COBRE.html). A total 103 subjects, including 52 patients and 51 controls were used.

The ADHD samples obtained from ADHD-200 Sample database, a rs-fMRI and anatomical MRI database pooled from eight independent imaging sites [36]. The subject data and phenotypic information can be downloaded from [http://fcon\\_1000.projects.nitrc.org/indi/adhd200/index.html](http://fcon_1000.projects.nitrc.org/indi/adhd200/index.html). The diagnosis criteria of each site can be obtained from the ADHD-200 Sample website. ADHD hyperactive-type subjects were excluded due to the small number of these subjects in the database. Therefore, we only classified the control subjects from two ADHD subtypes, the ADHD combined and ADHD inattentive types. The datasets from the NeuroIMAGE sample were removed because they were collected using a 1.5T system. The subjects from Washington University School of Medicine were also excluded because this dataset contained no ADHD subjects. After ruling out unacceptable subjects, a total 360 male subjects, including 189 patients and 171 controls, were used.

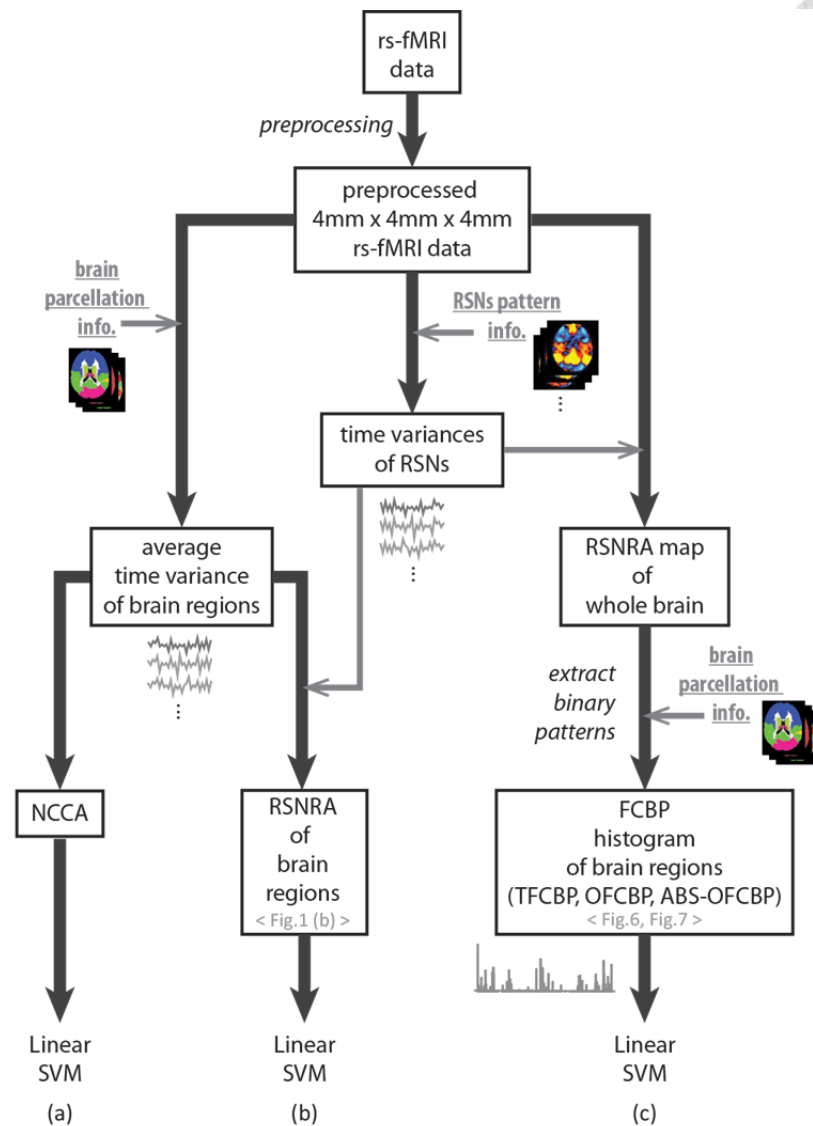
### ***ICA maps***

To show the generalization of the RSNs referenced approaches examined in this study, we directly applied the ICA results from previous studies [1]. These results can be downloaded from the FMRIB website (<http://fsl.fmrib.ox.ac.uk/analysis/brainmap+rsns/>). The ICA results, with twenty



components exhibiting the most correspondence between the connectivity information of CCA and ICA [181] and a reasonable resulting feature size were used. Two strategies were used to select the ICs in this study. In 20-RSNRA, all 20 ICs were used to describe the connectivity. Moreover, Smith et al. showed that 10 RSNs from the 20 ICs are well-matched to the fMRI activation networks [1]. The 10 RSNs might be the most informative elements to discriminate the mental disorders. In 10-RSNRA, we employed a simpler strategy, using only these 10 ICs.

Figure 6-4 shows the flow chart of all analyses used in this study, including NCCA, RSNRA, and FCBP. The details of each step are described below. We performed NCCA as the baseline for comparison with other methods. For training classifiers, the Pearson's correlation coefficients (CCs) for every paired brain region were used as features in NCCA.



**Figure 6-4** Flow chart of network cross-correlation analysis (NCCA), resting-state networks referenced analysis (RSNRA), and functional connectivity binary patterns (FCBP).

After preprocessing of rs-fMRI data, **(a)** the NCCA were performed by introducing brain parcellation information and computing the cross-correlation between each brain regions. **(b)** The time variances of resting state networks (RSNs) can be computed by using the introduced RSNs patterns. These time variances were used to perform the RSNRA of the brain regions or to build a whole brain RSNRA map. **(c)** The RSNRA map can be used to extract different binary patterns and build the FCBP histogram of brain regions, such as threshold (TFCBP), ordered (OFCBP), and absolute ordered (ABS-OFCBP) FCBP. The performance of each approach was evaluated through linear SVM and 10-fold cross-validation. The elements with bold font and underlined text show the introduced information.

### **Resting-state networks referenced analysis (RSNRA)**



To describe the connectivity of each voxel, NCCA was employed using the time series of all other voxels in the brain, resulting in a large number of correlations that were hard to implement. Fortunately, ICA provided a widely accepted framework to describe whole-brain connectivity using RSNs [1, 174]. Based on the assumption of ICA, all the rs-fMRI signals might be combined through ICs with different weights. As a stable and consistent subnetwork of whole-brain functional connectivity, RSNs from ICA might be adequate candidates for the references to describe the connectivity of each voxel in the brain [175]. Therefore, we used these RSNs as references and introduced the variations of these RSNs as simplified coordinates to describe the connectivity of each voxel.

Based on the ICA definition, for  $K$  independent components and  $J$  voxels in the data volume with  $t$  time points, the fMRI signal can be represented using the space-time data matrix:

$$X_{jt} = \sum_{k=1}^K A_{jk} S_{kt} + E_{jt} \quad (6-1)$$

where the columns of  $A$  denote the ICs, the rows of  $S$  represent the time-variation of each ICs, and  $E$  is the noise [171, 173, 176]. For the fMRI data of each subject,

$X_{jt}$  is known. Following our assumption, we consider the RSNs as the known prior,



and  $A_{jk}$  is also known. Therefore, we can use a simple linear regression to estimate the different  $S_{kt}$  of each network, that is, the variations of each RSN of this subject. Thus, the  $S_{kt}$  could be used as a references to describe the connectivity of each brain region (Figure 6-1 (b)). The method is referred to as the resting-state networks referenced analysis (RSNRA). Moreover, Figure 6-5 shows an example of using the ten RSNs to describe the connectivity of single voxel.

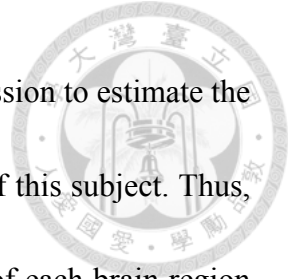
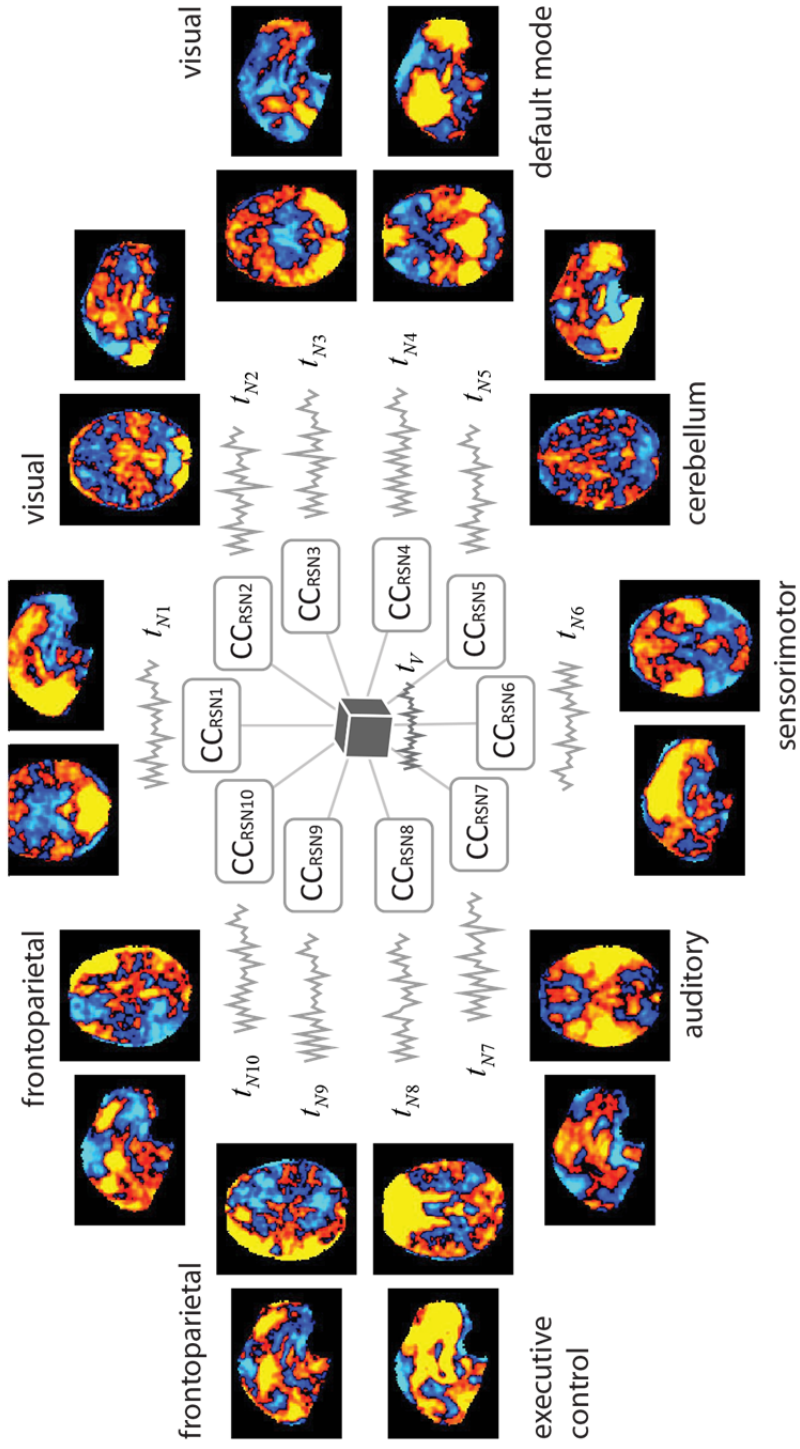


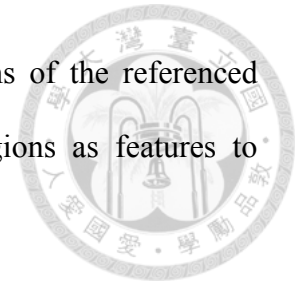
Figure 6-5 Connectivity of a single voxel described by ten resting-state networks (RSNs).



The referenced time series of N RSNs (in this case, the most informative ten RSNs in a previous study were used [1]) was established using the simple linear regression of the prior N RSN patterns and the rs-fMRI data. For describing the connectivity of each voxel in the brain, the Pearson's correlation coefficients of the voxel time course ( $t_v$ ) and the N reference RSNs time series ( $t_{N1} \sim t_{N10}$ ) are computed ( $CC_{RSN1} \sim CC_{RSN10}$ ). Then, the connectivity of each voxel can be described using these N values.



To evaluate the performance of RSNRA, we used the correlations of the referenced RSNs time course and the average time series of the brain regions as features to determine classification models of ADHD and schizophrenia.



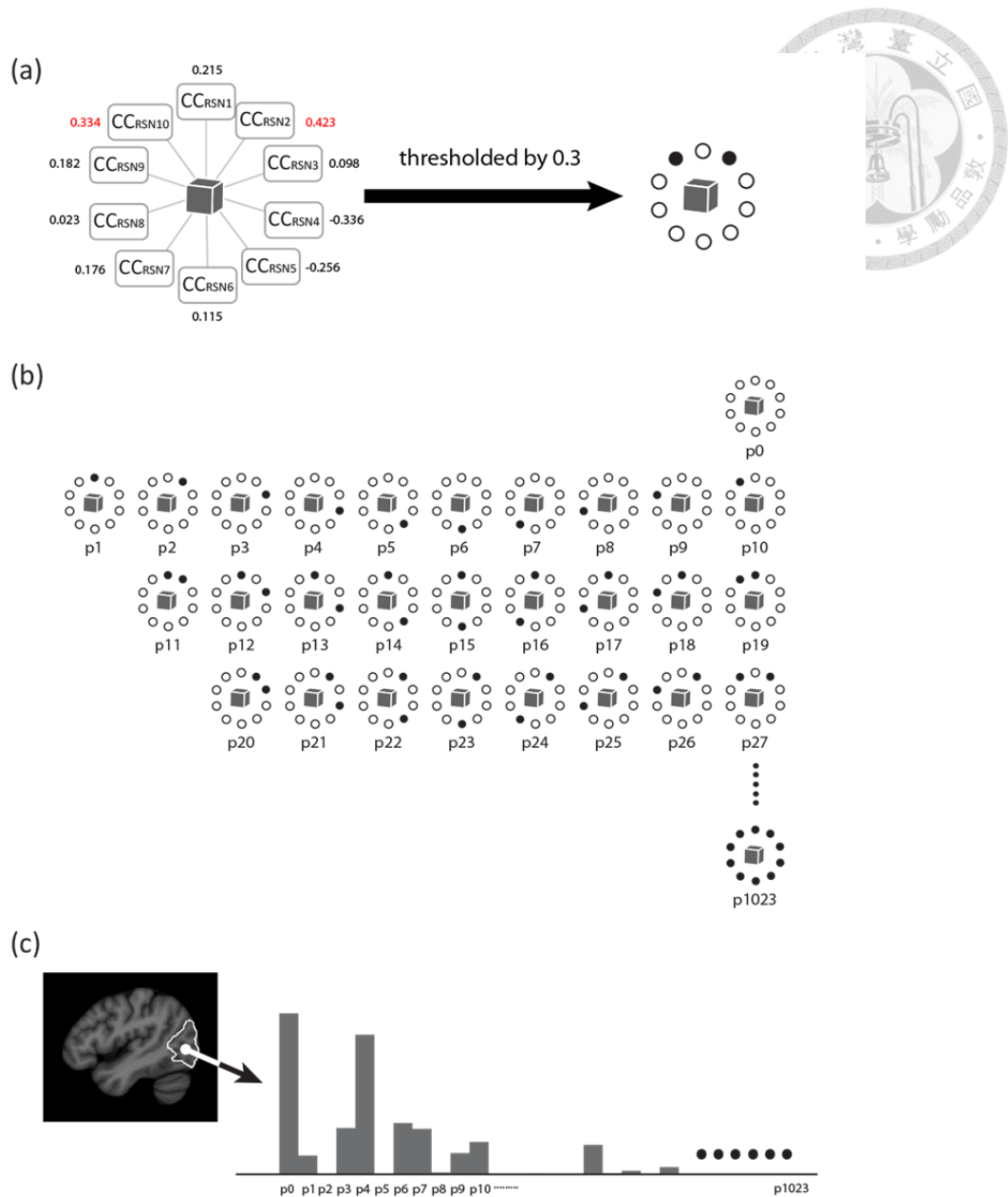
To combine the information in activation maps derived from BrainMap database, the RSNs from Smith et al. were directly used as the prior RSNs patterns [1]. The ICA results with twenty components were considered as the greatest correspondence between the connectivity information of CCA and ICA [181]. Therefore, the ICA results of twenty components would be used in this study. The ten RSNs in these twenty components contain the most informative RSNs because they are well-matched with the ICA results of 29,671-subject BrainMap activation database [1]. Both the RSNRA with a total of twenty RSNs time series (20-RSNRA) and the ten the most informative RSNs (10-RSNRA) were tested. To evaluate RSNRA, we compared the performance of NCCA, 20-RSNRA, and 10-RSNRA.

Based on the RSNRA strategy, this idea can be extended to the connectivity of any voxel of the brain (Figure 6-3 (b1) and Figure 6-5). Therefore, we changed our focus to the distribution constructed using the functional connectivity patterns in a specific brain region. In this section, we introduced two useful rules to transform the connectivity of each voxel to simple functional connectivity binary patterns (FCBP). Then, we established the distribution information defined by the statistics of the patterns (Figure 6-3 (b3)).

### ***Threshold functional connectivity binary patterns (TFCBP)***



First, as shown in Figure 6-6, an intuitional threshold method was introduced. Although the resulting feature size is large, the whole feature matrix is sparse. For example, when using 0.3 as the threshold, it is almost impossible to determine any voxel with more than three referenced CCs larger than 0.3. In this study, three possible threshold types were tested. We generated the threshold using negative CC, positive CC, and the combined histograms of negative CC and positive CC.



**Figure 6-6 Illustration of threshold functional connectivity binary patterns (TFCBP).**

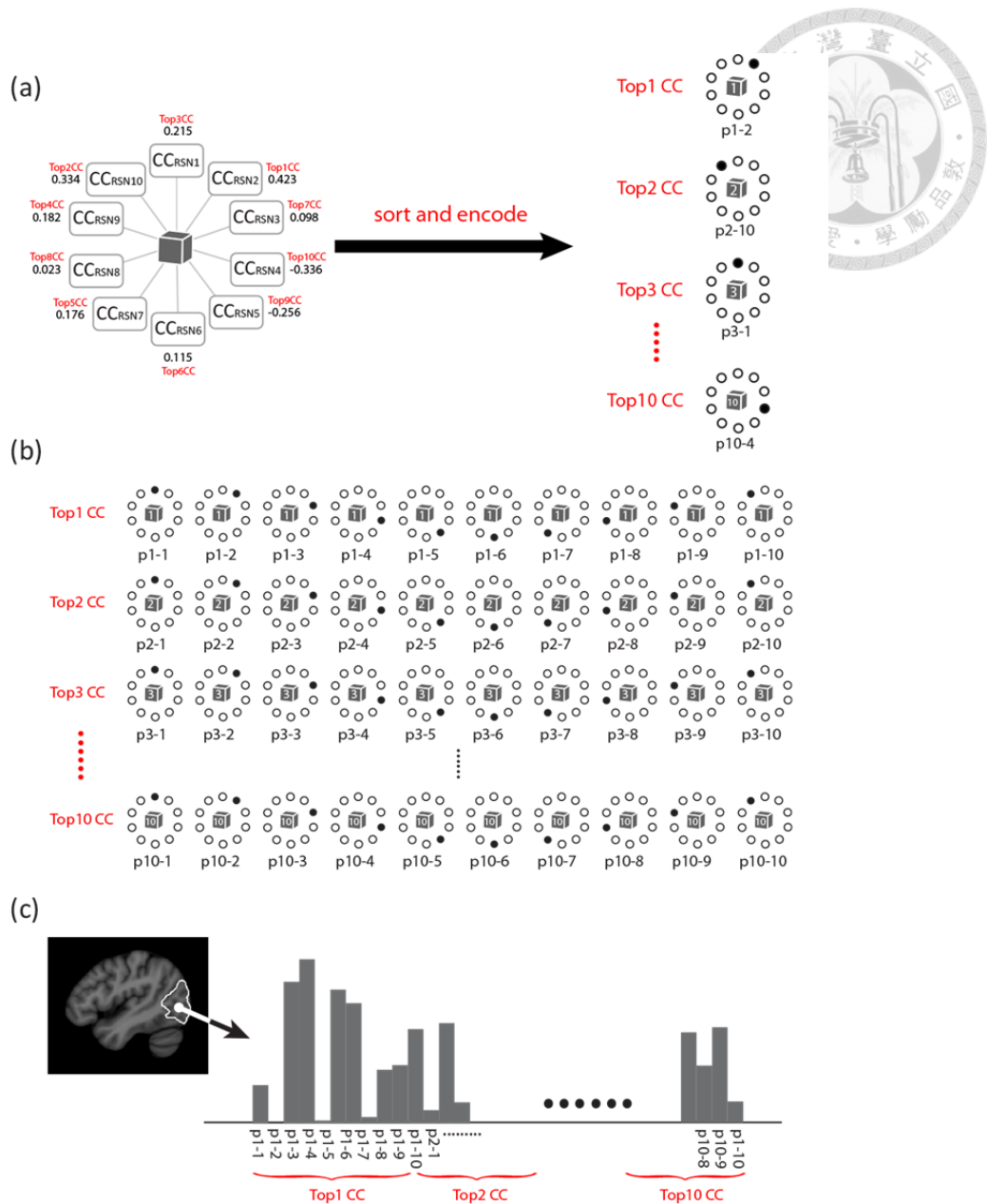
In TFCBP, a fixed threshold would be assigned to transform the Pearson's correlation coefficients between the time series of the voxel and the variances of referenced resting state networks ( $CC_{RSNs}$ ) to binary values. Figure (a) shows an example of using threshold 0.3 to build the binary pattern of specific voxel. In the right side of (a), the values larger than the threshold are shown as filled circles; otherwise this information is denoted as empty circles. (b) Considering all possible combinations, we will obtain  $2^k$  patterns of connectivity, where  $k$  is the number of used RSNs (in this case,  $k=10$ ).  $p_n$  shows the label of each distinct pattern. (c) The behaviors of one brain region would be described using  $2^k$  features, which is defined according to the histogram of these patterns.

### ***Ordered functional connectivity binary patterns (OFCBP)***



Two problems existed in the direct threshold binary patterns. First, although the resulting matrix is sparse, the raw feature size remained large. This size is difficult to use in multi-modality approaches. Second, identifying a proper threshold is not straightforward. Thus every possible threshold should be tested to obtain the best performance threshold.

Consequently, we provided another strategy to extract more abstractive connectivity information. We observed that the order of the correlations of RSNs might be the most informative aspect of these approaches. That is, the RSN with the most important role (with highest cross correlation) in the connectivity of a specific voxel could be most informative. Moreover, the RSN with second most important role should also be important. The order of the connectivity of each RSN of a specific voxel might provide more information for classification. Therefore, a simple binarization rule might be constructed based on this point of view. As shown in Figure 6-7, only the order of the connectivity was considered. As a result, only  $N^2$  features are considered.



**Figure 6-7 Illustration of ordered functional connectivity binary patterns (OFCBP).**

In OFCBP, only the order of Pearson's correlation coefficients (CC) between the time series of the voxel and the variances of reference resting state networks ( $CC_{RSNs}$ ) was considered. The RSN with largest CC (the Top1CC) is converted to the binary pattern p1-1 to p1-10. The RSN with top two CC (the Top2CC) is encoded to p2-1 to p2-10, and so on. Figure (a) shows an example of using OFCBP to build the binary patterns of specific voxel. In the right side of (a), the topNCC is encoded as a filled circle; otherwise, empty circles are used. The number of voxels indicates the top N RSN. Figure (b) shows part of all possible patterns. When using K RSNs,  $K^2$  features would be considered (in this case,  $k=10$ ). PN-M shows the Mst RSN in top N group. (c) The behaviors of one brain region would be described using  $K^2$  features, which are defined according to the histogram of the binary patterns.

### ***Absolute ordered functional connectivity binary patterns (ABS-OFCBP)***



To consider both the positive and negative CC in the connectivity, we also tested the order of original CC value and the order of the absolute CC value in OFCBP.

### ***Preprocessing of rs-fMRI data***

The preprocessing of each rs-fMRI data was primarily based on the Athena pipeline of the ADHD-200 preprocessed data website (<http://www.nitrc.org/plugins/mwiki/index.php/neurobureau:AthenaPipeline>). The preprocessing was performed using AFNI [150] and FSL [123]. After slice time and 3D motion corrections, each rs-fMRI dataset was transformed into NIHPD Objective 1 atlases (4.5~18.5y) [126, 127] with a 4 mm isotropic resolution. Then, the extracted WM and CSF time course and the motion time series were regressed out from the data. Subsequently, the data were blurred using a 6-mm FWHM Gaussian filter.

## **Results**

### ***Network cross-correlation analysis (NCCA)***

As shown in Table 6-1, using NCCA strategies, the best performance for classifying schizophrenia is 0.76 (Table 6-1, (a)). However, the classifier acquired almost no information from the ADHD database (Table 6-1, (b)). Comparisons between separated



parcellation sets using MNI and NCCA demonstrated the best performance in schizophrenia classification (Table 6-1, (a)). CORT provided the best accuracy in ADHD classification (Table 6-1, (b)).





**Table 6-1 Cross-validation results of resting-state networks referenced analysis (RSNRA) and network cross-correlation analysis (NCCA).**

(a) schizophrenia classification

parcellation		NCCA	10-RSNRA	20-RSNRA
	MNI	0.76	<u>0.72</u>	<u>0.72</u>
	CORT	0.68	0.69	0.68
	SUB-CORT	0.54	0.63	0.67
[a]	GM+WM	0.59	0.62	0.67
[b]	[a]+MNI	0.74	<u>0.65</u>	0.75
[c]	[b]+CORT	0.72	<u>0.69</u>	0.72
[d]	[c]+SUB-CORT	0.73	<u>0.73</u>	<u>0.71</u>

(b) ADHD classification

parcellation		NCCA	10-RSNRA	20-RSNRA
	MNI	0.506	0.536	0.503
	CORT	0.542	0.550	0.578
	SUB-CORT	0.508	0.519	0.539
[a]	GM+WM	0.528	0.544	0.536
[b]	[a]+MNI	0.519	0.556	<u>0.508</u>
[c]	[b]+CORT	0.550	<u>0.531</u>	0.561
[d]	[c]+SUB-CORT	0.544	0.547	0.556

The underlined results show the resulting accuracy is worse than the NCCA approaches using the same parcellation. The performance comparison of those approaches using McNemar's test can be found in Table 6-2.



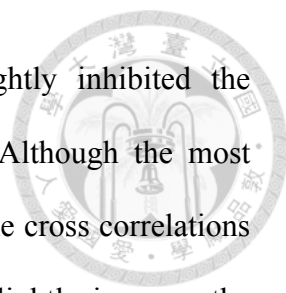
**Table 6-2 *p*-values of McNemar's test of results in Table 6-1.**  
 (a) schizophrenia classification

parcellation	10-RSNA vs NCCA	20-RSNRA vs NCCA
MNI	<u>0.5023</u>	<u>0.5224</u>
CORT	1.0000	1.0000
SUB-CORT	0.2432	0.0865
[a] GM+WM	0.7488	0.2913
[b] [a]+MNI	<u>0.0809</u>	1.0000
[c] [b]+CORT	<u>0.5791</u>	1.0000
[d] [c]+SUB-CORT	1.0000	<u>0.7518</u>

(b) ADHD classification

parcellation	10-RSNRA vs NCCA	20-RSNRA vs NCCA
MNI	0.3859	1.0000
CORT	0.8357	0.1770
SUB-CORT	0.8170	0.4188
[a] GM+WM	0.6790	0.8699
[b] [a]+MNI	0.2870	<u>0.7925</u>
[c] [b]+CORT	<u>0.5200</u>	0.7199
[d] [c]+SUB-CORT	1.0000	0.7373

No results with significant difference can be found while comparing the results of resting-state networks referenced analysis (RSNRA) and network cross-correlation analysis (NCCA). The underlined results show that the resulting accuracy is worse than the NCCA approaches using same parcellation.



Using merged parcellation sets, learning with fine regions slightly inhibited the performance when discriminating schizophrenia (Table 6-1, (a)). Although the most useful information for classifying ADHD might be obtained from the cross correlations of cortical regions, adding information from other regions could slightly improve the accuracy (Table 6-1, (b)).

#### ***Resting-state networks referenced analysis (RSNRA)***

In general, separated or merged parcellation sets, using RSNRA, provided equal performance with NCCA in our tests of both databases. Both 20-RSNRA and 10-RSNRA do not show significant differences with the NCCA results using McNemar's test (Table 6-2). 10-RSNRA showed an adequate performance in almost all cases. These results showed the ten most informative RSNs as candidates for describing the connectivity of specific brain regions or voxels. Therefore, we used these ten RSNs as reference RSNs in the remaining tests.

#### ***Threshold functional connectivity binary patterns (TFCBP)***

The results of TFCBP are shown in Table 6-3 and Table 6-4, and the results of McNemar's test are shown in Table 6-5. When classifying schizophrenia, the use of either a positive or negative CC threshold did not produce better results than those obtained using traditional NCCA approaches (Table 6-4). Combining the features of these approaches could improve the accuracy of the resulting model, exhibiting a more enhanced performance than NCCA. As shown in Table 6-3, the combined CC threshold



of 0.2 or 0.3 could improve the classification of schizophrenia. Despite the MNI parcellation, the performance of almost all parcellations with combined CC thresholds of 0.2 and 0.3 was better than that of NCCA approaches.

**Table 6-3 Cross-validation results of threshold functional connectivity binary patterns (TFCBP).**

(a) schizophrenia classification

		Combined CC Threshold				
		0.1	0.2	0.3	0.4	0.5
	MNI	<u>0.69</u>	<u>0.66</u>	<u>0.65</u>	** <u>0.56</u>	* <u>0.60</u>
	CORT	0.71	0.73	0.72	0.70	0.72
	SUB-CORT	0.69	0.69	* 0.70	0.64	0.62
[a]	GM+WM	0.68	0.70	0.70	0.63	0.70
[b]	[a]+MNI	<u>0.68</u>	<u>0.71</u>	<u>0.67</u>	* <u>0.60</u>	<u>0.64</u>
[c]	[b]+CORT	<u>0.71</u>	0.74	0.75	<u>0.69</u>	0.75
[d]	[c]+SUB-CORT	<u>0.71</u>	0.75	0.78	<u>0.69</u>	<u>0.66</u>

(b) ADHD classification

		Combined CC Threshold				
		0.1	0.2	0.3	0.4	0.5
	MNI	0.567	* 0.600	* 0.586	0.531	0.514
	CORT	* 0.611	0.606	* 0.622	* 0.622	0.569
	SUB-CORT	0.575	0.575	0.569	* 0.608	0.572
[a]	GM+WM	0.564	0.542	0.564	0.578	0.531
[b]	[a]+MNI	0.578	* 0.592	0.581	0.539	0.511
[c]	[b]+CORT	0.608	** 0.639	** 0.653	0.603	<u>0.558</u>
[d]	[c]+SUB-CORT	* 0.611	* 0.625	** 0.644	0.611	0.572

The underlined results show the resulting accuracy is worse than the network cross-correlation analysis (NCCA) approaches using same parcellation in Table 6-1. The stars show that the performance is significant different with NCCA approaches in McNemar's test (\*:  $p$ -value < 0.05. \*\*:  $p$ -value < 0.005). The details of each  $p$ -value in McNemar's test can be found in Table 6-5.

**Table 6-4 Cross-validation results of threshold functional connectivity binary patterns (TFCBP).**

		Negative CC Threshold					Positive CC Threshold					Combined CC Threshold				
		< -0.1	< -0.2	< -0.3	< -0.4	< -0.5	> 0.1	> 0.2	> 0.3	> 0.4	> 0.5	0.1	0.2	0.3	0.4	0.5
(a) schizophrenia classification																
	MNI	* <u>0.61</u>	<u>0.67</u>	<u>0.66</u>	* <u>0.63</u>	* <u>0.62</u>	<u>0.70</u>	<u>0.65</u>	<u>0.65</u>	<u>0.68</u>	<u>0.66</u>	<u>0.69</u>	<u>0.66</u>	<u>0.65</u>	** <u>0.56</u>	* <u>0.60</u>
	CORT	<u>0.66</u>	0.69	0.73	<u>0.62</u>	<u>0.62</u>	0.70	0.72	0.68	0.68	0.71	0.71	0.73	0.72	0.70	0.72
	SUB-CORT	** 0.73	* 0.69	0.65	0.59	0.62	0.69	0.56	0.63	0.61	0.64	0.69	0.69	* 0.70	0.64	0.62
[a]	GM+WM	0.65	<u>0.58</u>	0.61	0.60	0.71	0.65	0.63	0.60	0.63	<u>0.52</u>	0.68	0.70	0.70	0.63	0.70
[b]	[a]+MNI	<u>0.65</u>	<u>0.68</u>	<u>0.65</u>	<u>0.65</u>	<u>0.64</u>	<u>0.70</u>	<u>0.67</u>	<u>0.65</u>	<u>0.69</u>	<u>0.65</u>	<u>0.68</u>	<u>0.71</u>	<u>0.67</u>	* <u>0.60</u>	<u>0.64</u>
[c]	[b]+CORT	<u>0.71</u>	<u>0.70</u>	<u>0.71</u>	<u>0.62</u>	<u>0.63</u>	<u>0.67</u>	<u>0.70</u>	<u>0.67</u>	<u>0.69</u>	<u>0.71</u>	<u>0.71</u>	<u>0.74</u>	0.75	<u>0.69</u>	0.75
[d]	[c]+SUB-CORT	<u>0.71</u>	0.73	0.73	<u>0.64</u>	<u>0.61</u>	<u>0.70</u>	<u>0.72</u>	<u>0.70</u>	<u>0.66</u>	<u>0.70</u>	<u>0.71</u>	0.75	0.78	<u>0.69</u>	<u>0.66</u>
(b) ADHD classification																
	MNI	* <u>0.589</u>	* 0.581	0.558	** 0.617	** 0.606	* 0.581	0.572	0.550	* 0.578	0.542	0.567	* 0.600	* 0.586	0.531	0.514
	CORT	0.558	** 0.625	0.603	<u>0.522</u>	<u>0.542</u>	0.547	0.594	0.581	0.592	<u>0.531</u>	* 0.611	0.606	* 0.622	* 0.622	0.569
	SUB-CORT	0.575	0.542	0.508	0.519	0.553	* 0.597	* 0.583	0.528	<u>0.500</u>	<u>0.506</u>	0.575	0.575	* 0.569	* 0.608	0.572
[a]	GM+WM	0.561	0.586	* 0.603	0.592	0.522	0.581	0.575	0.600	0.572	0.486	0.564	0.542	0.564	0.578	0.531
[b]	[a]+MNI	* 0.603	0.575	0.578	* 0.606	0.589	0.567	0.575	0.553	0.575	<u>0.528</u>	0.578	* 0.592	0.581	0.539	0.511
[c]	[b]+CORT	0.606	0.603	* 0.619	0.547	0.544	0.592	0.603	0.597	0.597	<u>0.553</u>	0.608	** 0.639	** 0.653	0.603	<u>0.558</u>
[d]	[c]+SUB-CORT	0.581	0.606	* 0.619	0.589	0.553	0.586	0.586	* 0.611	0.597	0.542	* 0.611	* 0.625	** 0.644	0.611	0.572

The underlined results show that the resulting accuracy is worse than the network cross-correlation analysis (NCCA) approaches using the same parcellation as in Table 1. The stars show that the performance is significantly different from the NCCA approaches when assessed by McNemar's test (\*: p-value < 0.05. \*\*: p-value < 0.005). The details of each p-value in McNemar's test can be found in Table 6-5.

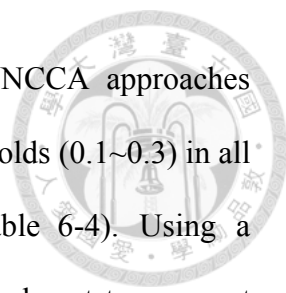


Table 6-5 p-values of McNemar's test of results in Table 6-3 and Table 6-4.

		Negative CC Threshold					Positive CC Threshold					Combined CC Threshold				
		< -0.1	< -0.2	< -0.3	< -0.4	< -0.5	> 0.1	> 0.2	> 0.3	> 0.4	> 0.5	0.1	0.2	0.3	0.4	0.5
(a) schizophrenia classification																
	MNI	<u><b>*0.0148</b></u>	0.1096	0.0550	<b>*0.0259</b>	<b>*0.0176</b>	<u>0.3447</u>	0.0725	0.0910	0.2433	0.1336	0.2482	0.0662	0.0543	<b>**0.0008</b>	<b>*0.0177</b>
	CORT	<u>0.8312</u>	1.0000	0.4414	0.3613	0.4047	0.8383	0.5403	1.0000	1.0000	0.7003	0.6625	0.4042	0.5224	0.8551	0.5403
	SUB-CORT	<b>**0.0087</b>	<b>*0.0455</b>	0.1273	0.5754	0.2801	0.0500	0.1344	0.2432	0.3602	0.2031	0.0545	0.0637	<b>*0.0375</b>	0.2291	0.3317
[a]	GM+WM	0.4510	1.0000	0.8676	1.0000	0.0744	0.4510	0.6434	1.0000	0.6583	0.3815	0.1884	0.1273	0.0910	0.6353	0.1360
[b]	[a]+MNI	0.1637	0.2864	0.0953	0.1096	0.1116	0.5839	0.2812	0.1763	0.5322	0.2330	0.3613	0.6464	0.2482	<b>*0.0176</b>	0.1336
[c]	[b]+CORT	1.0000	0.8026	1.0000	0.0890	0.1884	0.3827	0.8383	0.4237	0.6625	1.0000	1.0000	0.8137	0.6276	0.6625	0.6767
[d]	[c]+SUB-CORT	0.8026	1.0000	1.0000	0.1237	0.0820	0.6625	1.0000	0.6892	0.2482	0.6625	0.8137	0.8026	0.3320	0.5023	0.1904
(b) ADHD classification																
	MNI	<b>*0.0164</b>	<b>*0.0442</b>	0.1586	<b>**0.0028</b>	<b>**0.0083</b>	<b>*0.0274</b>	0.0708	0.2081	<b>*0.0495</b>	0.3413	0.0843	<b>*0.0095</b>	<b>*0.0264</b>	0.5543	0.8732
	CORT	0.6560	<b>**0.0098</b>	0.0780	0.6320	1.0000	0.9284	0.1186	0.2650	0.1419	0.8012	<b>*0.0278</b>	0.0509	<b>*0.0144</b>	<b>*0.0160</b>	0.4369
	SUB-CORT	0.0760	0.3932	1.0000	0.8159	0.2636	<b>*0.0161</b>	<b>*0.0442</b>	0.6574	0.8748	1.0000	0.0742	0.0708	0.1031	<b>*0.0066</b>	0.0925
[a]	GM+WM	0.4097	0.1284	<b>*0.0468</b>	0.0791	0.9358	0.1687	0.2211	0.0610	0.2662	0.3059	0.3531	0.7650	0.3671	0.1817	1.0000
[b]	[a]+MNI	<b>*0.0143</b>	0.1331	0.0921	<b>*0.0241</b>	0.0617	0.1685	0.1208	0.3456	0.1208	0.8732	0.0990	<b>*0.0481</b>	0.0969	0.6384	0.8724
[c]	[b]+CORT	0.0905	0.1213	<b>*0.0332</b>	1.0000	0.9358	0.2031	0.1074	0.1747	0.1685	1.0000	0.0599	<b>**0.0031</b>	<b>**0.0006</b>	0.1186	0.8699
[d]	[c]+SUB-CORT	0.2753	0.0780	<b>*0.0181</b>	0.2298	0.8724	0.2105	0.2105	<b>*0.0486</b>	0.1158	1.0000	<b>*0.0312</b>	<b>*0.0116</b>	<b>**0.0008</b>	0.0502	0.4501



The underlined results show that the resulting accuracy is worse than the network cross-correlation analysis (NCCA) approaches using the same parcellation as in Table 1. The stars show that the performance is significantly different from the NCCA approaches when assessed by McNemar's test (\*: p-value < 0.05. \*\*: p-value < 0.005).



The results of discriminating ADHD are more interesting. As NCCA approaches acquired almost nothing from the rs-fMRI data, all reasonable thresholds (0.1~0.3) in all parcellation sets demonstrated better accuracy than NCCA (Table 6-4). Using a combined CC threshold dramatically increased the performance to almost ten percent better than NCCA approaches.

***Ordered functional connectivity binary patterns (OFCBP)***

As shown in Table 6-6 and Table 6-5, the classifications of both disorders provide better performance than NCCA. However, these performances are slightly worse than the performance using TFCBP. In the ADHD classification, the results of models using MNI, [b], and [d] parcellations are significantly better than those in NCCA.

Nevertheless, the classification of schizophrenia showed a different pattern, as models with MNI and [b] performed worse than those with NCCA, and only the model with SUB-CORT was significantly better than those with NCCA.





**Table 6-6 Cross-validation results of ordered functional connectivity binary patterns (OFCBP and absolute OFCBP, ABS-OFCBP) and network cross-correlation analysis (NCCA).**

(a) schizophrenia classification

parcellation		NCCA	OFCBP	ABS-OFCBP
	MNI	0.76	<u>0.71</u>	0.77
	CORT	0.68	0.73	0.76
	SUB-CORT	0.54	* 0.71	* 0.70
[a]	GM+WM	0.59	0.69	0.65
[b]	[a]+MNI	0.74	<u>0.67</u>	0.75
[c]	[b]+CORT	0.72	0.74	0.78
[d]	[c]+SUB-CORT	0.73	0.74	0.78

(b) ADHD classification

parcellation		NCCA	OFCBP	ABS-OFCBP
	MNI	0.506	** 0.603	** 0.625
	CORT	0.542	0.578	0.578
	SUB-CORT	0.508	0.542	0.561
[a]	GM+WM	0.528	0.614	0.564
[b]	[a]+MNI	0.519	** 0.581	** 0.628
[c]	[b]+CORT	0.550	0.608	0.589
[d]	[c]+SUB-CORT	0.544	* 0.625	0.592

The underlined results show that the resulting accuracy is worse than the NCCA approaches using the same parcellation as in Table 6-1. The stars show that the performance is significantly different from the NCCA approaches when assessed by McNemar's test (\*:  $p$ -value < 0.05. \*\*:  $p$ -value < 0.005). The details of each  $p$ -value in McNemar's test can be found in Table 6-7.



**Table 6-7 p-values of McNemar's test of results in Table 6-6.**  
 (a) schizophrenia classification

parcellation		OFCBP vs NCCA	ABS-OFCBP vs NCCA
	MNI	0.3588	1.0000
	CORT	0.4042	0.1530
	SUB-CORT	* 0.0223	* 0.0375
[a]	GM+WM	0.1649	0.4173
[b]	[a]+MNI	<u>0.1456</u>	1.0000
[c]	[b]+CORT	0.8137	0.1814
[d]	[c]+SUB-CORT	1.0000	0.2278

(b) ADHD classification

parcellation		OFCBP vs NCCA	ABS-OFCBP vs NCCA
	MNI	** 0.0048	** 0.0011
	CORT	0.2370	0.2753
	SUB-CORT	0.3010	0.1761
[a]	GM+WM	0.0906	0.3643
[b]	[a]+MNI	** 0.0050	** 0.0017
[c]	[b]+CORT	0.3552	0.2274
[d]	[c]+SUB-CORT	* 0.0420	0.1254

The underlined results show that the resulting accuracy is worse than the network cross-correlation analysis (NCCA) approaches using the same parcellation as in Table 6-1. The stars show that the performance is significantly different from the NCCA approaches when assessed by McNemar's test (\*:  $p$ -value < 0.05. \*\*:  $p$ -value < 0.005).

### ***Absolute ordered binary pattern distribution (ABS-OFCBP)***



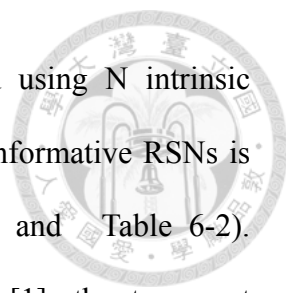
Again, these two disorders showed different patterns in this test (Table 6-6 and Table 6-7). Using ABS-OFCBP increased the performance in almost all cases for the classification of schizophrenia. However, the effects of sorting using absolute CC are not consistent in ADHD cases. ABS-OFCBP improved the performance of the models with MNI, SUB-CORT, and [b]. However, this process negatively affected the accuracy of the original OFCBP approaches.

### **Discussion**

#### ***Resting-state networks referenced analysis (RSNRA)***

In this study, we showed that the RSNRA is a simpler alternative to NCCA for classifying schizophrenia and ADHD. The most obvious advantage of using RSNRA is the marked reduction of feature size when using many regions (Figure 6-1). Using RSNRA instead of NCCA largely reduces the feature size of the resting-state brain connectivity with the same performance (Table 6-1 and Table 6-2). As a simpler method to explore complex whole-brain functional connectivity, RSNRA could be a good candidate for use in multi-modality studies, combined with other approaches.

Moreover, RSNRA converts the complex whole-brain functional connectivity graph to a simple matrix with  $M \times N$  dimensions, in which  $M$  is the considered regions, and  $N$  is the



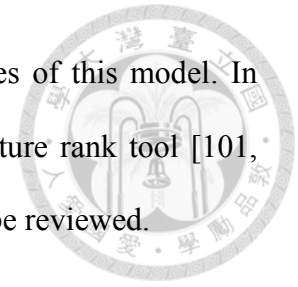
selected RSNs. The characters of each region can be described using  $N$  intrinsic reference networks. These results suggest that using the ten most informative RSNs is sufficient to discriminate schizophrenia and ADHD (Table 6-1 and Table 6-2). Moreover, if we combine the RSNs used in a previous study [1], the ten most informative RSNs can be associated with the knowledge in the large fMRI studies database, BrainMap [183, 184], as demonstrated in Figure 6-8.

### ***Binary pattern distribution***

Three FCBP distributions have been tested in this study, TFCBP, OFCBP, and ABS-OFCBP. Using distribution patterns to classify both disorders provides better performance than traditional NCCA (Table 6-3 and Table 6-6). In simplified ordered distribution patterns, ABS-OFCBP was best suited for schizophrenia, and OFCBP exhibited better performance in classifying ADHD (Table 6-6). In TFCBP approaches, the best threshold to use still needs to be identified, but several interesting results were observed.

First, both positive and negative information is useful for classifying both disorders (Table 6-3). Combining negative and positive features generated the best performance. In most traditional rs-fMRI studies, only positive connectivity was considered. However, the results of this study showed the importance of the information revealed by the negative connectivity.

Second, using linear SVM approaches, we can explore the features of this model. In addition to being a good classifier, linear-SVM is also a good feature rank tool [101, 103]. Thus, the content of the resulting classification model should be reviewed.



### ***Knowledge discovered in classification models***

Figure 6-8 shows the top 100 features of the classification model of both disorders using TFCBP, with a 0.3 threshold, combining both positive and negative features. Although the use of these features alone cannot provide sufficient classifiers, an analysis of top N features might reveal information concerning how the model was built and provide insight for future studies. The two disorders show different patterns.

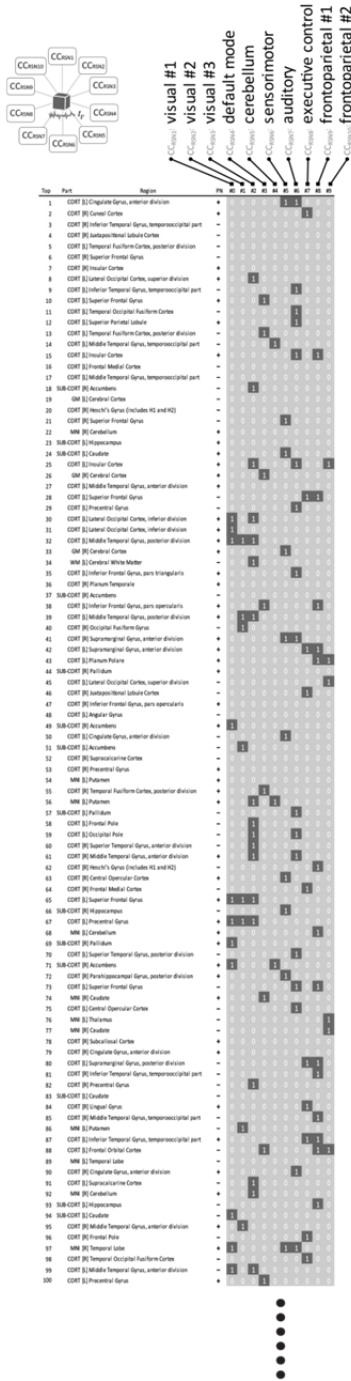
(a) accumulation number of each RSN in top 100 features of schizophrenia classification model

RSN	acc. number
RSN3	visual#3 18
RSN7	auditory 17
RSN9	frontoparietal#1 14
RSN1	visual#1 11
RSN8	executive control 10
RSN6	sensorimotor 10
RSN2	visual#2 8
RSN4	default mode 8
RSN10	frontoparietal#2 6
RSN5	cerebellum 3

(b) accumulation number of each RSN in top 100 features of ADHD classification model

RSN	acc. number
RSN8	executive control 20
RSN7	auditory 18
RSN6	sensorimotor 9
RSN9	frontoparietal#1 9
RSN3	visual#3 7
RSN1	visual#1 6
RSN4	default mode 6
RSN10	frontoparietal#2 5
RSN2	visual#2 3
RSN5	cerebellum 1

(c) schizophrenia classification



(d) ADHD classification

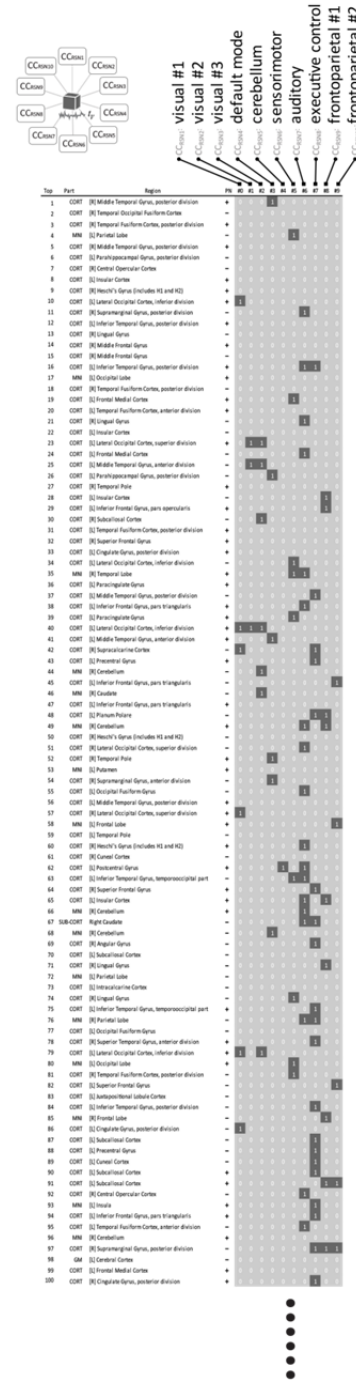


Figure 6-8 Top 100 features in the classification models of schizophrenia and ADHD using threshold functional connectivity binary patterns with 0.3 CC thresholds and combining positive and negative features.

(a) and (b) show the accumulation number of each RSN in the top 100 features. (c) and (d) show the details of top 100 features. Each row in (c) and (d) shows the pattern presented by

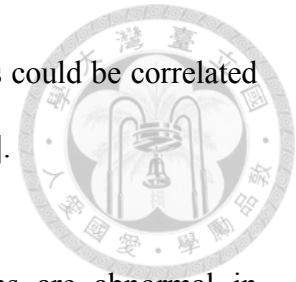
the feature. The column "Top" shows the rank of the feature. The columns "Part" and "Region" show the brain region the pattern extracted from. The column "PN" shows the positive or the negative threshold the feature used. Columns #0 to #9 show the binary patterns of the feature. As shown in Figure 6-6, a value of 1 indicates the correlation coefficient value between the voxel time series and the reference time series is larger/lower than the threshold, otherwise, the value would be 0.

The results shown in Figure 6-8 (c) and Figure 6-8 (d) could be analogous to a DNA-microarray analysis. Each cell denotes the binary connection activity of specific RSN and brain region. Various features represent different combinations of the activation patterns of those cells. Therefore, disparate mental disorders show distinct patterns of features.

For both disorders, most of the top 20 features are defined according to the number of voxels with "no link" patterns (no correlation with any RSNs) in several brain regions (Figure 6-8 (c) and Figure 6-8 (d)), potentially suggesting no connectivity with any RSNs; that is, no activity in the resting-state. Therefore, the proportion of no active voxels might represent an over-all resting-state activity measurement in the corresponding brain regions. The activity of these regions might be different between patients and controls.

Intuitively, one could analyze the brain regions represented in the top 100 features. More interestingly, using the approaches proposed in the present study, we could view these data as highlighting the roles of RSNs in each disorder. Evaluating the number of RSNs in the top 100 features will reveal the importance of each RSN in classifying the

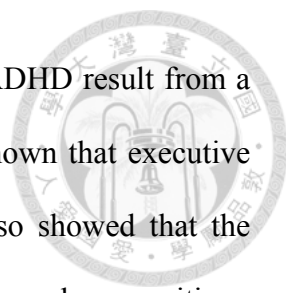
disorder (Figure 6-8 (a) and Figure 6-8 (b)). Moreover, these results could be correlated with the event-related fMRI results, as shown in a previous study [1].



For the schizophrenia model (Figure 6-8 (c)), many top regions are abnormal in schizophrenia patients, as evidenced through anatomical or functional studies, such as Inferior Temporal Gyrus (top 3) [185], Supplementary Motor Cortex (Juxtapositional Lobule Cortex, top 4) [186, 187], Temporal Fusiform Cortex (top 5) [188], Superior Frontal Gyrus (top 6) [189], and Insular Cortex (top 7) [190]. From the perspective of RSN, as shown in Fig. 8(a), the distribution associated with many RSNs, such as the auditory network, visual networks, and frontoparietal#1 network, could discriminate schizophrenia. The importance of the auditory and visual networks might reflect major symptoms of schizophrenia, such as delusions and auditory hallucinations [191]. To date, most rs-fMRI studies on schizophrenia have focused on the default mode network (DMN) [12]; however, the results obtained in the present study suggested that DMN might not play an important role in classifying this disorder. Thus, studying other networks might reveal more information.

For the ADHD model (Figure 6-8 (d)), several top regions, such as Temporal Occipital Fusiform Cortex (top 2) [192], Temporal Fusiform Cortex (top 3) [193], Middle Temporal Gyrus (top 5) [194], Parahippocampal Gyrus (top 6) [194], and Insular Cortex (top 8) [195], have also shown differences between ADHD and controls. However, the executive control and auditory networks are the most important networks for classifying ADHD (Figure 6-8 (b)). Therefore, it is reasonable to conclude that executive control is the most important network for classifying ADHD. Several well-known





neuropsychological theories have suggested that the symptoms of ADHD result from a primary deficit in executive functions [196]. Much evidence has shown that executive functions play an important role in ADHD [197]. Smith et al. also showed that the auditory network strongly corresponds to action–execution–speech, cognition–language–speech, and perception–audition paradigms[1]. This importance might reflect the non-stop talking symptoms of hyperactivity.

There are two contributions in this study. First, we used resting-state networks referenced analysis (RSNRA) to convert the complex NCCA to a much simpler representation using time series of RSNs as references and showed the equivalent performance of these RSNs. Second, we proposed using TFCBP and OFCBP to view the rs-fMRI data in order to analyze the distribution of connectivity. Our results showed that the proposed method consistently improved the accuracy of classification for schizophrenia and significantly enhanced the performance when discriminating ADHD (Table 6-3 and Table 6-6). Moreover, the resulting model could provide information about these disorders for future studies. This study introduced another perspective from which to analyze the rs-fMRI data using the distribution of FCBP.

## 6.3 Conclusion

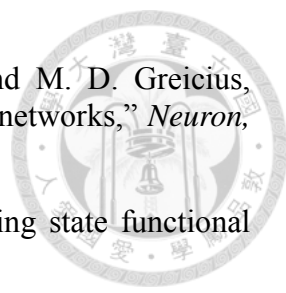
In this work, we tried to apply simple binary patterns methods on both brain structural and functional MRI data to extract useful information. Results showed those simple binary methods are useful for extract information from structural and functional brain MR images. Those methods are good candidates to be used in large-scale brain associated big data researches.

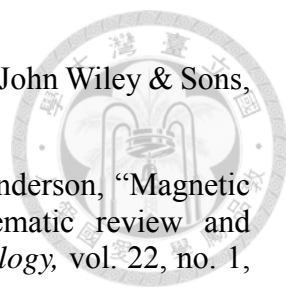


## REFERENCE

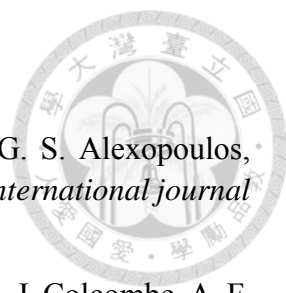


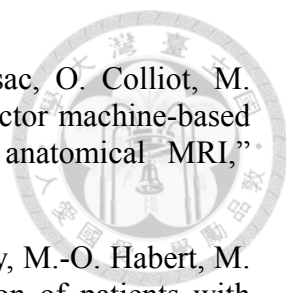
- [1] S. M. Smith, P. T. Fox, K. L. Miller, D. C. Glahn, P. M. Fox, C. E. Mackay, N. Filippini, K. E. Watkins, R. Toro, and A. R. Laird, "Correspondence of the brain's functional architecture during activation and rest," *Proceedings of the National Academy of Sciences*, vol. 106, no. 31, pp. 13040-13045, 2009.
- [2] A. Halevy, P. Norvig, and F. Pereira, "The unreasonable effectiveness of data," *Intelligent Systems, IEEE*, vol. 24, no. 2, pp. 8-12, 2009.
- [3] M. Richardson, "A funding profile of the NIH," *Research Trends*, no. 34, 2013.
- [4] L. L. Wald, "The future of acquisition speed, coverage, sensitivity, and resolution," *Neuroimage*, 2012.
- [5] OECD, *Health at a glance 2011: OECD indicators*, Paris: Organization for Economic Cooperation and Development, 2011.
- [6] M. Chung, I. J. Dahabreh, N. Hadar, S. J. Ratichek, J. M. Gaylor, T. A. Trikalinos, and J. Lau, "Emerging MRI Technologies for Imaging Musculoskeletal Disorders Under Loading Stress," 2011.
- [7] A. W. Toga, and J. C. Mazziotta, *Brain mapping: the systems*: Academic Press, San Diego, 2000.
- [8] P. J. Basser, J. Mattiello, and D. LeBihan, "MR diffusion tensor spectroscopy and imaging," *Biophysical journal*, vol. 66, no. 1, pp. 259-267, 1994.
- [9] Y. Assaf, and O. Pasternak, "Diffusion tensor imaging (DTI)-based white matter mapping in brain research: a review," *Journal of Molecular Neuroscience*, vol. 34, no. 1, pp. 51-61, 2008.
- [10] J. S. Damoiseaux, and M. D. Greicius, "Greater than the sum of its parts: a review of studies combining structural connectivity and resting-state functional connectivity," *Brain Structure and Function*, vol. 213, no. 6, pp. 525-533, 2009.
- [11] B. Biswal, F. Zerrin Yetkin, V. M. Haughton, and J. S. Hyde, "Functional connectivity in the motor cortex of resting human brain using echo-planar mri," *Magnetic Resonance in Medicine*, vol. 34, no. 4, pp. 537-541, 1995.
- [12] C. Rosazza, and L. Minati, "Resting-state brain networks: literature review and clinical applications," *Neurological sciences*, vol. 32, no. 5, pp. 773-785, 2011.
- [13] M. Lee, C. Smyser, and J. Shimony, "Resting-State fMRI: A Review of Methods and Clinical Applications," *American Journal of Neuroradiology*, 2012.

- 
- [14] W. W. Seeley, R. K. Crawford, J. Zhou, B. L. Miller, and M. D. Greicius, "Neurodegenerative diseases target large-scale human brain networks," *Neuron*, vol. 62, no. 1, pp. 42, 2009.
- [15] M. D. Fox, and M. Greicius, "Clinical applications of resting state functional connectivity," *Front Syst Neurosci*, vol. 4, 2010.
- [16] M. D. Greicius, G. Srivastava, A. L. Reiss, and V. Menon, "Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 13, pp. 4637-4642, 2004.
- [17] C. Sorg, V. Riedl, M. Mühlau, V. D. Calhoun, T. Eichele, L. Läer, A. Drzezga, H. Förstl, A. Kurz, and C. Zimmer, "Selective changes of resting-state networks in individuals at risk for Alzheimer's disease," *Proceedings of the National Academy of Sciences*, vol. 104, no. 47, pp. 18760-18765, 2007.
- [18] Y. I. Sheline, and M. E. Raichle, "Resting State Functional Connectivity in Preclinical Alzheimer's Disease," *Biol Psychiatry*, 2013.
- [19] M. Rocca, P. Valsasina, M. Absinta, G. Riccitelli, M. Rodegher, P. Misci, P. Rossi, A. Falini, G. Comi, and M. Filippi, "Default-mode network dysfunction and cognitive impairment in progressive MS," *Neurology*, vol. 74, no. 16, pp. 1252-1259, 2010.
- [20] S. Bonavita, A. Gallo, R. Sacco, M. Della Corte, A. Bisecco, R. Docimo, L. Lavorgna, D. Corbo, A. Di Costanzo, and F. Tortora, "Distributed changes in default-mode resting-state connectivity in multiple sclerosis," *Multiple Sclerosis Journal*, vol. 17, no. 4, pp. 411-422, 2011.
- [21] L. Breiman, "Statistical modeling: The two cultures (with comments and a rejoinder by the author)," *Statistical Science*, vol. 16, no. 3, pp. 199-231, 2001.
- [22] V. Mayer-Schönberger, and K. Cukier, *Big Data: A Revolution that Will Transform how We Live, Work, and Think*: Eamon Dolan/Houghton Mifflin Harcourt, 2013.
- [23] J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant, "Detecting influenza epidemics using search engine query data," *Nature*, vol. 457, no. 7232, pp. 1012-1014, 2008.
- [24] H. Jin, Q. Liu, H. Lu, and X. Tong, "Face detection using improved LBP under bayesian framework." pp. 306-309.
- [25] C. N. E. Anagnostopoulos, I. E. Anagnostopoulos, V. Loumos, and E. Kayafas, "A license plate-recognition algorithm for intelligent transportation system applications," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 7, no. 3, pp. 377-392, 2006.

- 
- [26] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*: John Wiley & Sons, 2012.
- [27] D. Arnone, A. McIntosh, K. Ebmeier, M. Munafò, and I. Anderson, “Magnetic resonance imaging studies in unipolar depression: systematic review and meta-regression analyses,” *European Neuropsychopharmacology*, vol. 22, no. 1, pp. 1-16, 2012.
- [28] A. Etkin, and T. D. Wager, “Functional neuroimaging of anxiety: a meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia,” *The American journal of psychiatry*, vol. 164, no. 10, pp. 1476, 2007.
- [29] K. K. Zakzanis, S. J. Graham, and Z. Campbell, “A meta-analysis of structural and functional brain imaging in dementia of the Alzheimer's type: a neuroimaging profile,” *Neuropsychology Review*, vol. 13, no. 1, pp. 1-18, 2003.
- [30] I. Ellison-Wright, and E. Bullmore, “Anatomy of bipolar disorder and schizophrenia: a meta-analysis,” *Schizophrenia research*, vol. 117, no. 1, pp. 1-12, 2010.
- [31] G. Orrù, W. Pettersson-Yeo, A. F. Marquand, G. Sartori, and A. Mechelli, “Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review,” *Neuroscience & Biobehavioral Reviews*, vol. 36, no. 4, pp. 1140-1152, 2012.
- [32] J.-B. Poline, J. L. Breeze, S. Ghosh, K. Gorgolewski, Y. O. Halchenko, M. Hanke, C. Haselgrove, K. G. Helmer, D. B. Keator, and D. S. Marcus, “Data sharing in neuroimaging research,” *Frontiers in neuroinformatics*, vol. 6, 2012.
- [33] R. P. Guralnick, A. W. Hill, and M. Lane, “Towards a collaborative, global infrastructure for biodiversity assessment,” *Ecology letters*, vol. 10, no. 8, pp. 663-672, 2007.
- [34] T. A. Manolio, L. D. Brooks, and F. S. Collins, “A HapMap harvest of insights into the genetics of common disease,” *The Journal of clinical investigation*, vol. 118, no. 5, pp. 1590, 2008.
- [35] K. P. Andriole, J. M. Wolfe, R. Khorasani, S. T. Treves, D. J. Getty, F. L. Jacobson, M. L. Steigner, J. J. Pan, A. Sitek, and S. E. Seltzer, “Optimizing analysis, visualization, and navigation of large image data sets: One 5000-section ct scan can ruin your whole day,” *Radiology*, vol. 259, no. 2, pp. 346-362, 2011.
- [36] M. P. Milham, D. Fair, M. Mennes, and S. H. Mostofsky, “The ADHD-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience,” *Front Syst Neurosci*, vol. 6, pp. 62, 2012.
- [37] N. Raz, and K. M. Rodrigue, “Differential aging of the brain: patterns, cognitive correlates and modifiers,” *Neuroscience & Biobehavioral Reviews*, vol. 30, no. 6,

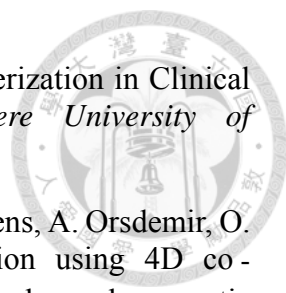
pp. 730-748, 2006.

- 
- [38] F. M. Gunning-Dixon, A. M. Brickman, J. C. Cheng, and G. S. Alexopoulos, "Aging of cerebral white matter: a review of MRI findings," *International journal of geriatric psychiatry*, vol. 24, no. 2, pp. 109-117, 2009.
- [39] K. M. Kennedy, K. I. Erickson, K. M. Rodrigue, M. W. Voss, S. J. Colcombe, A. F. Kramer, J. D. Acker, and N. Raz, "Age-related differences in regional brain volumes: a comparison of optimized voxel-based morphometry to manual volumetry," *Neurobiology of Aging*, vol. 30, no. 10, pp. 1657-1676, 2009.
- [40] T. J. Passe, P. Rajagopalan, L. A. Tupler, C. E. Byrum, J. R. Macfall, and K. Krishnan, "Age and sex effects on brain morphology," *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, vol. 21, no. 8, pp. 1231-1237, 1997.
- [41] C. D. Good, I. S. Johnsrude, J. Ashburner, R. N. Henson, K. Fristen, and R. S. Frackowiak, "A voxel-based morphometric study of ageing in 465 normal adult human brains." p. 16 pp.
- [42] N. Gogtay, J. N. Giedd, L. Lusk, K. M. Hayashi, D. Greenstein, A. C. Vaituzis, T. F. Nugent, D. H. Herman, L. S. Clasen, and A. W. Toga, "Dynamic mapping of human cortical development during childhood through early adulthood," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 21, pp. 8174-8179, 2004.
- [43] V. A. Kovalev, F. Kruggel, and D. Y. Von Cramon, "Gender and age effects in structural brain asymmetry as measured by MRI texture analysis," *Neuroimage*, vol. 19, no. 3, pp. 895-905, 2003.
- [44] E. R. Sowell, B. S. Peterson, P. M. Thompson, S. E. Welcome, A. L. Henkenius, and A. W. Toga, "Mapping cortical change across the human life span," *Nature neuroscience*, vol. 6, no. 3, pp. 309-315, 2003.
- [45] Z. Lao, D. Shen, Z. Xue, B. Karacali, S. M. Resnick, and C. Davatzikos, "Morphological classification of brains via high-dimensional shape transformations and machine learning methods," *Neuroimage*, vol. 21, no. 1, pp. 46-57, 2004.
- [46] Y. Fan, D. Shen, R. C. Gur, R. E. Gur, and C. Davatzikos, "COMPARE: classification of morphological patterns using adaptive regional elements," *Medical Imaging, IEEE Transactions on*, vol. 26, no. 1, pp. 93-105, 2007.
- [47] S. Klöppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Scahill, J. D. Rohrer, N. C. Fox, C. R. Jack, J. Ashburner, and R. S. J. Frackowiak, "Automatic classification of MR scans in Alzheimer's disease," *Brain*, vol. 131, no. 3, pp. 681-689, 2008.


- 
- [48] B. Magnin, L. Mesrob, S. Kinkingnehun, M. Péligrini-Issac, O. Colliot, M. Sarazin, B. Dubois, S. Lehericy, and H. Benali, “Support vector machine-based classification of Alzheimer’s disease from whole-brain anatomical MRI,” *Neuroradiology*, vol. 51, no. 2, pp. 73-83, 2009.
- [49] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehericy, M.-O. Habert, M. Chupin, H. Benali, and O. Colliot, “Automatic classification of patients with Alzheimer’s disease from structural MRI: A comparison of ten methods using the ADNI database,” *Neuroimage*, vol. 56, no. 2, pp. 766-781, 2011.
- [50] A. Mechelli, C. J. Price, K. J. Friston, and J. Ashburner, “Voxel-based morphometry of the human brain: methods and applications,” *Current Medical Imaging Reviews*, vol. 1, no. 2, pp. 105-113, 2005.
- [51] I. Wright, P. McGuire, J.-B. Poline, J. Traverso, R. Murray, C. Frith, R. Frackowiak, and K. Friston, “A voxel-based method for the statistical analysis of gray and white matter density applied to schizophrenia,” *Neuroimage*, vol. 2, no. 4, pp. 244-252, 1995.
- [52] J. Ashburner, and K. J. Friston, “Voxel-based morphometry—the methods,” *Neuroimage*, vol. 11, no. 6, pp. 805-821, 2000.
- [53] P. A. Bandettini, “What’s new in neuroimaging methods?,” *Ann N Y Acad Sci*, vol. 1156, pp. 260-93, Mar, 2009.
- [54] C. J. Mummary, K. Patterson, C. Price, J. Ashburner, R. Frackowiak, and J. R. Hodges, “A voxel-based morphometry study of semantic dementia: relationship between temporal lobe atrophy and semantic memory,” *Annals of neurology*, vol. 47, no. 1, pp. 36-45, 2000.
- [55] T. R. Franklin, P. D. Acton, J. A. Maldjian, J. D. Gray, J. R. Croft, C. A. Dackis, C. P. O’Brien, and A. R. Childress, “Decreased gray matter concentration in the insular, orbitofrontal, cingulate, and temporal cortices of cocaine patients,” *Biological psychiatry*, vol. 51, no. 2, pp. 134-142, 2002.
- [56] E. J. Burton, I. G. McKeith, D. J. Burn, E. D. Williams, and J. T. O’Brien, “Cerebral atrophy in Parkinson’s disease with and without dementia: a comparison with Alzheimer’s disease, dementia with Lewy bodies and controls,” *Brain*, vol. 127, no. 4, pp. 791-800, 2004.
- [57] N. Focke, G. Helms, S. Kaspar, C. Diederich, V. Tóth, P. Dechent, A. Mohr, and W. Paulus, “Multi-site voxel-based morphometry—Not quite there yet,” *Neuroimage*, vol. 56, no. 3, pp. 1164-1170, 2011.
- [58] F. L. Bookstein, ““Voxel-based morphometry” should not be used with imperfectly registered images,” *Neuroimage*, vol. 14, no. 6, pp. 1454-1462, 2001.
- [59] W. Crum, L. Griffin, D. Hill, and D. Hawkes, “Zen and the art of medical image

- registration: correspondence, homology, and quality,” *Neuroimage*, vol. 20, no. 3, pp. 1425-1437, 2003.
- [60] A. Klein, J. Andersson, B. A. Ardekani, J. Ashburner, B. Avants, M. C. Chiang, G. E. Christensen, D. L. Collins, J. Gee, P. Hellier, J. H. Song, M. Jenkinson, C. Lepage, D. Rueckert, P. Thompson, T. Vercauteren, R. P. Woods, J. J. Mann, and R. V. Parsey, “Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration,” *Neuroimage*, vol. 46, no. 3, pp. 786-802, Jul 1, 2009.
- [61] H. B. M. Uylings, G. Rajkowska, E. Sanz-Arigita, K. Amunts, and K. Zilles, “Consequences of large interindividual variability for human brain atlases: converging macroscopical imaging and microscopical neuroanatomy,” *Anat Embryol (Berl)*, vol. 210, no. 5, pp. 423-431, 2005.
- [62] M. Mirmehdi, X. Xie, and J. Suri, *Handbook of texture analysis*: Imperial College Press, 2009.
- [63] S. Herlidou-Meme, J. Constans, B. Carsin, D. Olivie, P. Eliat, L. Nadal-Desbarats, C. Gondry, E. Le Rumeur, I. Idy-Peretti, and J. de Certaines, “MRI texture analysis on texture test objects, normal brain and intracranial tumors,” *Magn Reson Imaging*, vol. 21, no. 9, pp. 989-993, 2003.
- [64] E. I. Zacharaki, S. Wang, S. Chawla, D. Soo Yoo, R. Wolf, E. R. Melhem, and C. Davatzikos, “Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme,” *Magnetic Resonance in Medicine*, vol. 62, no. 6, pp. 1609-1618, 2009.
- [65] T. Sankar, N. Bernasconi, H. Kim, and A. Bernasconi, “Temporal lobe epilepsy: Differential pattern of damage in temporopolar cortex and white matter,” *Hum Brain Mapp*, vol. 29, no. 8, pp. 931-944, 2008.
- [66] M. De Oliveira, M. Balthazar, A. D'abreu, C. Yasuda, B. Damasceno, F. Cendes, and G. Castellano, “MR imaging texture analysis of the corpus callosum and thalamus in amnesic mild cognitive impairment and mild Alzheimer disease,” *American Journal of Neuroradiology*, vol. 32, no. 1, pp. 60-66, 2011.
- [67] Y. Zhang, H. Zhu, J. Mitchell, F. Costello, and L. M. Metz, “T2 MRI texture analysis is a sensitive measure of tissue injury and recovery resulting from acute inflammatory lesions in multiple sclerosis,” *Neuroimage*, vol. 47, no. 1, pp. 107, 2009.
- [68] M. C. Alegro, A. V. Silva, S. Y. Bando, R. D. Lopes, L. H. Martins de Castro, W. HungTsu, C. A. Moreira-Filho, and E. Amaro Jr, “Texture analysis of high resolution MRI allows discrimination between febrile and afebrile initial precipitating injury in mesial temporal sclerosis,” *Magnetic Resonance in Medicine*, 2012.

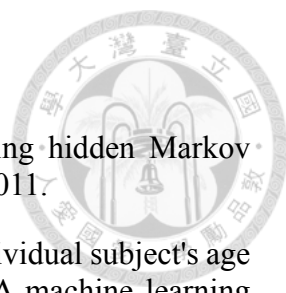


- 
- [69] K. K. Holli, "Texture Analysis as a Tool for Tissue Characterization in Clinical MRI," *Tampereen teknillinen yliopisto. Julkaisu-Tampere University of Technology. Publication;* 988, 2011.
- [70] B. J. Woods, B. D. Clymer, T. Kurc, J. T. Heverhagen, R. Stevens, A. Orsdemir, O. Bulan, and M. V. Knopp, "Malignant - lesion segmentation using 4D co - occurrence texture analysis applied to dynamic contrast-enhanced magnetic resonance breast image data," *Journal of Magnetic Resonance Imaging*, vol. 25, no. 3, pp. 495-501, 2007.
- [71] S. C. Agner, S. Soman, E. Libfeld, M. McDonald, K. Thomas, S. Englander, M. A. Rosen, D. Chin, J. Noshier, and A. Madabhushi, "Textural kinetics: A novel dynamic contrast-enhanced (dce)-mri feature for breast lesion classification," *Journal of Digital Imaging*, vol. 24, no. 3, pp. 446-463, 2011.
- [72] T. Mäenpää, and M. Pietikäinen, "Texture analysis with local binary patterns."
- [73] M. P. Inen, M. Pietikäinen, A. Hadid, G. Zhao, and T. Ahonen, *Computer Vision Using Local Binary Patterns*: Springer Verlag, 2011.
- [74] M. Varma, and A. Zisserman, "A statistical approach to material classification using image patch exemplars," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 11, pp. 2032-2047, 2009.
- [75] T. Ojala, M. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions." pp. 582-585.
- [76] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition*, vol. 29, no. 1, pp. 51-59, 1996.
- [77] T. Ojala, and M. Pietikäinen, "Unsupervised texture segmentation using feature distributions," *Pattern Recognition*, vol. 32, no. 3, pp. 477-486, 1999.
- [78] M. Pietikäinen, T. Ojala, and Z. Xu, "Rotation-invariant texture classification using feature distributions," *Pattern Recognition*, vol. 33, no. 1, pp. 43-52, 2000.
- [79] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Gray scale and rotation invariant texture classification with local binary patterns," *Computer Vision-ECCV 2000*, pp. 404-420: Springer, 2000.
- [80] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971-987, 2002.
- [81] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary

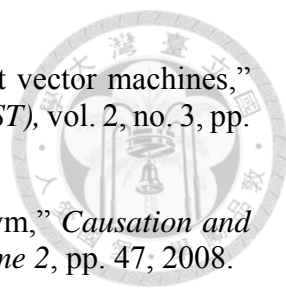
- patterns," *Computer Vision-ECCV 2004*, pp. 469-481: Springer, 2004.
- [82] M. Heikkila, and M. Pietikainen, "A texture-based method for modeling the background and detecting moving objects," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 4, pp. 657-662, 2006.
- [83] G. Zhao, and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 6, pp. 915-928, 2007.
- [84] A. Hadid, and M. Pietikäinen, "Combining appearance and motion for face and gender recognition from videos," *Pattern Recognition*, vol. 42, no. 11, pp. 2818-2827, 2009.
- [85] V. Kellokumpu, G. Zhao, and M. Pietikäinen, "Human activity recognition using a dynamic texture based method." pp. 1-10.
- [86] V. Kellokumpu, G. Zhao, and M. Pietikäinen, "Recognition of human actions using texture descriptors," *Machine Vision and Applications*, vol. 22, no. 5, pp. 767-780, 2011.
- [87] V. Kellokumpu, G. Zhao, S. Z. Li, and M. Pietikäinen, "Dynamic texture based gait recognition," *Advances in Biometrics*, pp. 1000-1009: Springer, 2009.
- [88] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 12, pp. 2037-2041, 2006.
- [89] A. Hadid, M. Pietikainen, and T. Ahonen, "A discriminative feature space for detecting and recognizing faces." pp. 797-804.
- [90] B. A. Rosdi, C. W. Shing, and S. A. Suandi, "Finger vein recognition using local line binary pattern," *Sensors*, vol. 11, no. 12, pp. 11357-11371, 2011.
- [91] T. Tommasi, and F. Orabona, "Idiap on Medical Image Classification ImageCLEF," The Information Retrieval Series H. Müller, P. Clough, T. Deselaers and B. Caputo, eds., pp. 453-465: Springer Berlin Heidelberg, 2010.
- [92] D. Unay, A. Ekin, and R. Jasinschi, "Medical image search and retrieval using local binary patterns and KLT feature points." pp. 997-1000.
- [93] D. Unay, A. Ekin, M. Cetin, R. Jasinschi, and A. Ercil, "Robustness of local binary patterns in brain MR image analysis." pp. 2098-2101.
- [94] D. Unay, and A. Ekin, "Intensity versus texture for medical image search and retrieval." pp. 241-244.
- [95] X. Gao, Y. Qian, M. Loomes, R. Comley, B. Barn, A. Chapman, J. Rix, R. Hui,

- 
- and Z. Tian, "Retrieval of 3D Medical Images via Their Texture Features," *International Journal On Advances in Software*, vol. 4, no. 3 and 4, pp. 499-509, 2012.
- [96] Y. Qian, X. Gao, M. Loomes, R. Comley, B. Barn, R. Hui, and Z. Tian, "Content-based retrieval of 3D medical images." pp. 7-12.
- [97] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers." pp. 144-152.
- [98] R. Herbrich, *Learning kernel classifiers: theory and algorithms*: Mit Press, 2002.
- [99] B. Schölkopf, and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization and beyond*: the MIT Press, 2002.
- [100] R. E. Fan, K. W. Chang, C. J. Hsieh, X. R. Wang, and C. J. Lin, "LIBLINEAR: A library for large linear classification," *The Journal of Machine Learning Research*, vol. 9, pp. 1871-1874, 2008.
- [101] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine learning*, vol. 46, no. 1-3, pp. 389-422, 2002.
- [102] D. Mladenić, J. Brank, M. Grobelnik, and N. Milic-Frayling, "Feature selection using linear classifier weights: interaction with classification models." pp. 234-241.
- [103] Y.-W. Chang, and C.-J. Lin, "Feature ranking using linear SVM," *Causation and Prediction Challenge Challenges in Machine Learning, Volume 2*, pp. 47, 2008.
- [104] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," 2003.
- [105] M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, and S. M. Smith, "FSL," *Neuroimage*, vol. 62, no. 2, pp. 782-790, 2012.
- [106] J. Davis, and M. Goadrich, "The relationship between Precision-Recall and ROC curves." pp. 233-240.
- [107] K. Franke, E. Luders, A. May, M. Wilke, and C. Gaser, "Brain maturation: Predicting individual BrainAGE in children and adolescents using structural MRI," *Neuroimage*, 2012.
- [108] K. Franke, G. Ziegler, S. Klöppel, and C. Gaser, "Estimating the age of healthy subjects from T<sub>1</sub>-weighted MRI scans using kernel methods: Exploring the influence of various parameters," *Neuroimage*, vol. 50, no. 3, pp. 883-892, 2010.
- [109] B. Wang, and T. D. Pham, "HMM-based brain age interpolation using kriging

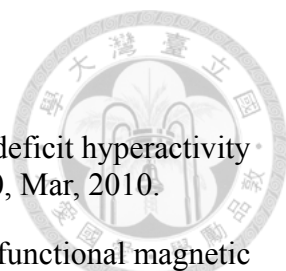
estimator." pp. 704-708.

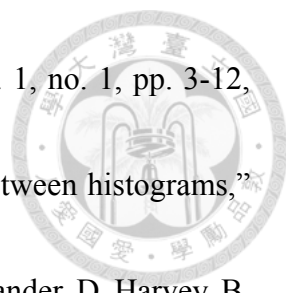
- 
- [110] B. Wang, and T. D. Pham, "MRI-based age prediction using hidden Markov models," *J Neurosci Methods*, vol. 199, no. 1, pp. 140-145, 2011.
- [111] B. Mwangi, K. M. Hasan, and J. C. Soares, "Prediction of individual subject's age across the human lifespan using diffusion tensor imaging: A machine learning approach," *Neuroimage*, 2013.
- [112] L. Su, L. Wang, F. Chen, H. Shen, B. Li, and D. Hu, "Sparse representation of brain aging: extracting covariance patterns from structural MRI," *PLoS One*, vol. 7, no. 5, pp. e36147, 2012.
- [113] J. Acosta-Cabronero, G. B. Williams, J. Pereira, G. Pengas, and P. J. Nestor, "The impact of skull-stripping and radio-frequency bias correction on grey-matter segmentation for voxel-based morphometry," *Neuroimage*, vol. 39, no. 4, pp. 1654-1665, 2008.
- [114] S. A. Sadananthan, W. Zheng, M. W. L. Chee, and V. Zagorodnov, "Skull stripping using graph cuts," *Neuroimage*, vol. 37, no. 1, pp. 225, 2010.
- [115] J. E. Iglesias, C. Y. Liu, P. M. Thompson, and Z. Tu, "Robust brain extraction across datasets and comparison with publicly available methods," *Medical Imaging, IEEE Transactions on*, vol. 30, no. 9, pp. 1617-1634, 2011.
- [116] C. Lebel, L. Walker, A. Leemans, L. Phillips, and C. Beaulieu, "Microstructural maturation of the human brain from childhood to adulthood," *Neuroimage*, vol. 40, no. 3, pp. 1044-1055, 2008.
- [117] E. V. Sullivan, and A. Pfefferbaum, "Diffusion tensor imaging and aging," *Neuroscience & Biobehavioral Reviews*, vol. 30, no. 6, pp. 749-761, 2006.
- [118] H. Zhang, P. A. Yushkevich, D. C. Alexander, and J. C. Gee, "Deformable registration of diffusion tensor MR images with explicit orientation optimization," *Med Image Anal*, vol. 10, no. 5, pp. 764-785, 2006.
- [119] P.-T. Yap, G. Wu, H. Zhu, W. Lin, and D. Shen, "F-TIMER: fast tensor image morphing for elastic registration," *Medical Imaging, IEEE Transactions on*, vol. 29, no. 5, pp. 1192-1203, 2010.
- [120] A. Gupta, M. Escolar, C. Dietrich, J. Gilmore, G. Gerig, and M. Styner, "3D tensor normalization for improved accuracy in DTI tensor registration methods," *Biomedical Image Registration*, pp. 170-179: Springer, 2012.
- [121] V. J. Wedeen, D. L. Rosene, R. Wang, G. Dai, F. Mortazavi, P. Hagmann, J. H. Kaas, and W.-Y. I. Tseng, "The geometric structure of the brain fiber pathways," *Science*, vol. 335, no. 6076, pp. 1628-1634, 2012.
- [122] C. W. Chang, C. C. Ho, and J. H. Chen, "ADHD classification by a texture

- analysis of anatomical brain MRI data,” *Front Syst Neurosci*, vol. 6, pp. 66, 2012.
- [123] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. J. Behrens, H. Johansen-Berg, P. R. Bannister, M. De Luca, I. Drobnjak, and D. E. Flitney, “Advances in functional and structural MR image analysis and implementation as FSL,” *Neuroimage*, vol. 23, pp. S208-S219, 2004.
- [124] Y. Zhang, M. Brady, and S. Smith, “Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm,” *Medical Imaging, IEEE Transactions on*, vol. 20, no. 1, pp. 45-57, 2001.
- [125] S. M. Smith, “Fast robust automated brain extraction,” *Hum Brain Mapp*, vol. 17, no. 3, pp. 143-55, Nov, 2002.
- [126] V. Fonov, A. Evans, R. McKinstry, C. Almlı, and D. Collins, “Unbiased nonlinear average age-appropriate brain templates from birth to adulthood,” *Neuroimage*, vol. 47, pp. S102, 2009.
- [127] V. Fonov, A. C. Evans, K. Botteron, C. R. Almlı, R. C. McKinstry, and D. L. Collins, “Unbiased average age-appropriate atlases for pediatric studies,” *Neuroimage*, vol. 54, no. 1, pp. 313, 2011.
- [128] J. Mazziotta, A. Toga, A. Evans, P. Fox, J. Lancaster, K. Zilles, R. Woods, T. Paus, G. Simpson, and B. Pike, “A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM),” *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 356, no. 1412, pp. 1293-1322, 2001.
- [129] J. Diedrichsen, J. H. Balsters, J. Flavell, E. Cussans, and N. Ramnani, “A probabilistic MR atlas of the human cerebellum,” *Neuroimage*, vol. 46, no. 1, pp. 39-46, 2009.
- [130] M. Jenkinson, P. Bannister, M. Brady, and S. Smith, “Improved optimization for the robust and accurate linear registration and motion correction of brain images,” *Neuroimage*, vol. 17, no. 2, pp. 825-841, 2002.
- [131] B. A. Ardekani, S. Guckemus, A. Bachman, M. J. Hoptman, M. Wojtaszek, and J. Nierenberg, “Quantitative comparison of algorithms for inter-subject registration of 3D volumetric brain MRI scans,” *J Neurosci Methods*, vol. 142, no. 1, pp. 67-76, Mar 15, 2005.
- [132] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers." pp. 144-152.
- [133] V. Vapnik, *The nature of statistical learning theory*: springer, 1999.
- [134] K. P. Bennett, and C. Campbell, “Support vector machines: hype or hallelujah?,” *ACM SIGKDD Explorations Newsletter*, vol. 2, no. 2, pp. 1-13, 2000.

- 
- [135] C.-C. Chang, and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, pp. 27, 2011.
- [136] Y. W. Chang, and C. J. Lin, "Feature ranking using linear svm," *Causation and Prediction Challenge Challenges in Machine Learning, Volume 2*, pp. 47, 2008.
- [137] J. Brank, M. Grobelnik, N. Milic-Frayling, and D. Mladenic, "Feature selection using linear support vector machines."
- [138] M. Cabezas, A. Oliver, X. Lladó, J. Freixenet, and M. Bach Cuadra, "A review of atlas-based segmentation for magnetic resonance brain images," *Comput Methods Programs Biomed*, vol. 104, no. 3, pp. e158-e177, 2011.
- [139] T. T. Brown, J. M. Kuperman, Y. Chung, M. Erhart, C. McCabe, D. J. Hagler, V. K. Venkatraman, N. Akshoomoff, D. G. Amaral, and C. S. Bloss, "Neuroanatomical assessment of biological maturity," *Current Biology*, 2012.
- [140] J. Ashburner, "A fast diffeomorphic image registration algorithm," *Neuroimage*, vol. 38, no. 1, pp. 95-113, 2007.
- [141] C. Lustig, P. Shah, R. Seidler, and P. A. Reuter-Lorenz, "Aging, training, and the brain: a review and future directions," *Neuropsychology review*, vol. 19, no. 4, pp. 504-522, 2009.
- [142] N. Raz, U. Lindenberger, K. M. Rodrigue, K. M. Kennedy, D. Head, A. Williamson, C. Dahle, D. Gerstorf, and J. D. Acker, "Regional brain changes in aging healthy adults: general trends, individual differences and modifiers," *Cerebral Cortex*, vol. 15, no. 11, pp. 1676-1689, 2005.
- [143] T. M. Pham, B. Winblad, A.-C. Granholm, and A. H. Mohammed, "Environmental influences on brain neurotrophins in rats," *Pharmacology Biochemistry and Behavior*, vol. 73, no. 1, pp. 167-175, 2002.
- [144] R. K. Lenroot, and J. N. Giedd, "Brain development in children and adolescents: insights from anatomical magnetic resonance imaging," *Neuroscience and biobehavioral reviews*, vol. 30, no. 6, pp. 718-729, 2006.
- [145] Y. Fu, G. Guo, and T. S. Huang, "Age synthesis and estimation via faces: A survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 11, pp. 1955-1976, 2010.
- [146] A. Lanitis, "Facial Age Estimation," *Scholarpedia*, vol. 5, no. 1, pp. 9701, 2010.
- [147] !!! INVALID CITATION !!!
- [148] J. H. Ferguson, "National Institutes of Health Consensus Development Conference Statement: Diagnosis and treatment of attention-deficit/hyperactivity disorder (ADHD)," *Journal of the American Academy of Child & Adolescent*

*Psychiatry*, vol. 39, no. 2, pp. 182-193, 2000.

- 
- [149] S. V. Faraone, and E. Mick, “Molecular genetics of attention deficit hyperactivity disorder,” *Psychiatr Clin North Am*, vol. 33, no. 1, pp. 159-80, Mar, 2010.
- [150] R. W. Cox, “AFNI: software for analysis and visualization of functional magnetic resonance neuroimages,” *Comput Biomed Res*, vol. 29, no. 3, pp. 162-73, Jun, 1996.
- [151] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, “Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain,” *Neuroimage*, vol. 15, no. 1, pp. 273-89, Jan, 2002.
- [152] R. C. Craddock, G. A. James, P. E. Holtzheimer, 3rd, X. P. Hu, and H. S. Mayberg, “A whole brain fMRI atlas generated via spatially constrained spectral clustering,” *Hum Brain Mapp*, Jul 18, 2011.
- [153] M. Ojala, and G. C. Garriga, “Permutation tests for studying classifier performance,” *The Journal of Machine Learning Research*, vol. 11, pp. 1833-1863, 2010.
- [154] P. I. Good, *Permutation tests*, 2nd edition ed., New York: Springer, 2000.
- [155] T. G. Dietterich, “Approximate statistical tests for comparing supervised classification learning algorithms,” *Neural computation*, vol. 10, no. 7, pp. 1895-1923, 1998.
- [156] B. Everitt, *The analysis of contingency tables*, second edition ed., London: Chapman & Hall/CRC, 1992.
- [157] R. Team, “R: a language and environment for statistical computing. 2007,” *R Foundation for Statistical Computing: Vienna, Austria*, 2012.
- [158] I. Ellison-Wright, Z. Ellison-Wright, and E. Bullmore, “Structural brain change in Attention Deficit Hyperactivity Disorder identified by meta-analysis,” *BMC Psychiatry*, vol. 8, pp. 51, 2008.
- [159] E. M. Valera, S. V. Faraone, K. E. Murray, and L. J. Seidman, “Meta-analysis of structural imaging findings in attention-deficit/hyperactivity disorder,” *Biol Psychiatry*, vol. 61, no. 12, pp. 1361-9, Jun 15, 2007.
- [160] S. M. Wolosin, M. E. Richardson, J. G. Hennessey, M. B. Denckla, and S. H. Mostofsky, “Abnormal cerebral cortex structure in children with ADHD,” *Hum Brain Mapp*, vol. 30, no. 1, pp. 175-84, Jan, 2009.
- [161] V. D. Calhoun, T. Eichele, and G. Pearlson, “Functional brain networks in schizophrenia: a review,” *Frontiers in Human Neuroscience*, vol. 3, 2009.

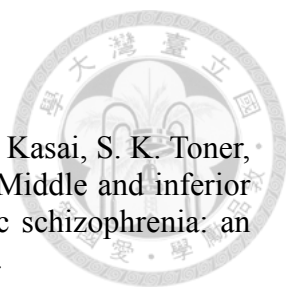
- 
- [162] M. E. Raichle, "The restless brain," *Brain Connectivity*, vol. 1, no. 1, pp. 3-12, 2011.
- [163] S. H. Cha, and S. N. Srihari, "On measuring the distance between histograms," *Pattern recognition*, vol. 35, no. 6, pp. 1355-1370, 2002.
- [164] C. R. Jack, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, and C. Ward, "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *Journal of Magnetic Resonance Imaging*, vol. 27, no. 4, pp. 685-691, 2008.
- [165] H. Yang, Q.-Z. Wu, L.-T. Guo, Q.-Q. Li, X.-Y. Long, X.-Q. Huang, R. C. Chan, and Q.-Y. Gong, "Abnormal spontaneous brain activity in medication-naïve ADHD children: A resting state fMRI study," *Neuroscience letters*, vol. 502, no. 2, pp. 89-93, 2011.
- [166] L. Ma, B. Wang, X. Chen, and J. Xiong, "Detecting functional connectivity in the resting brain: a comparison between ICA and CCA," *Magn Reson Imaging*, vol. 25, no. 1, pp. 47-56, 2007.
- [167] E. Bullmore, and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature Reviews Neuroscience*, vol. 10, no. 3, pp. 186-198, 2009.
- [168] N. U. F. Dosenbach, B. Nardos, A. L. Cohen, D. A. Fair, J. D. Power, J. A. Church, S. M. Nelson, G. S. Wig, A. C. Vogel, and C. N. Lessov-Schlaggar, "Prediction of individual brain maturity using fMRI," *Science*, vol. 329, no. 5997, pp. 1358-1361, 2010.
- [169] J. S. Anderson, J. A. Nielsen, A. L. Froehlich, M. B. DuBray, T. J. Druzgal, A. N. Cariello, J. R. Cooperrider, B. A. Zielinski, C. Ravichandran, and P. T. Fletcher, "Functional connectivity magnetic resonance imaging classification of autism," *Brain*, vol. 134, no. 12, pp. 3742-3754, 2011.
- [170] G. Chen, B. D. Ward, C. Xie, W. Li, Z. Wu, J. L. Jones, M. Franczak, P. Antuono, and S.-J. Li, "Classification of Alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional MR imaging," *Radiology*, vol. 259, no. 1, pp. 213-221, 2011.
- [171] 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," *Hum Brain Mapp*, vol. 6, pp. 160-188, 1998.
- [172] V. Kiviniemi, J. H. Kantola, J. Jauhiainen, A. Hyvärinen, and O. Tervonen, "Independent component analysis of nondeterministic fMRI signal sources," *Neuroimage*, vol. 19, no. 2, pp. 253-260, 2003.
- [173] C. F. Beckmann, and S. M. Smith, "Probabilistic independent component analysis

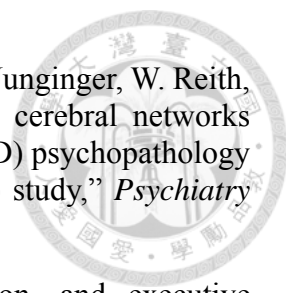


for functional magnetic resonance imaging,” *Medical Imaging, IEEE Transactions on*, vol. 23, no. 2, pp. 137-152, 2004.

- [174] C. F. Beckmann, M. DeLuca, J. T. Devlin, and S. M. Smith, “Investigations into resting-state connectivity using independent component analysis,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 360, no. 1457, pp. 1001-1013, 2005.
- [175] J. Damoiseaux, S. Rombouts, F. Barkhof, P. Scheltens, C. Stam, S. M. Smith, and C. Beckmann, “Consistent resting-state networks across healthy subjects,” *Proceedings of the National Academy of Sciences*, vol. 103, no. 37, pp. 13848-13853, 2006.
- [176] M. J. McKeown, L. K. Hansen, and T. J. Sejnowski, “Independent component analysis of functional MRI: what is signal and what is noise?,” *Current opinion in neurobiology*, vol. 13, no. 5, pp. 620, 2003.
- [177] D. M. Cole, S. M. Smith, and C. F. Beckmann, “Advances and pitfalls in the analysis and interpretation of resting-state FMRI data,” *Front Syst Neurosci*, vol. 4, 2010.
- [178] X.-Y. Long, X.-N. Zuo, V. Kiviniemi, Y. Yang, Q.-H. Zou, C.-Z. Zhu, T.-Z. Jiang, H. Yang, Q.-Y. Gong, and L. Wang, “Default mode network as revealed with multiple methods for resting-state functional MRI analysis,” *J Neurosci Methods*, vol. 171, no. 2, pp. 349-355, 2008.
- [179] M. P. van den Heuvel, and H. E. Hulshoff Pol, “Exploring the brain network: a review on resting-state fMRI functional connectivity,” *European Neuropsychopharmacology*, vol. 20, no. 8, pp. 519-534, 2010.
- [180] K. R. A. Van Dijk, T. Hedden, A. Venkataraman, K. C. Evans, S. W. Lazar, and R. L. Buckner, “Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization,” *Journal of neurophysiology*, vol. 103, no. 1, pp. 297-321, 2010.
- [181] C. Rosazza, L. Minati, F. Ghielmetti, M. Mandelli, and M. Bruzzone, “Functional Connectivity during Resting-State Functional MR Imaging: Study of the Correspondence between Independent Component Analysis and Region-of-Interest-Based Methods,” *American Journal of Neuroradiology*, vol. 33, no. 1, pp. 180-187, 2012.
- [182] M. Mennes, B. Biswal, F. X. Castellanos, and M. P. Milham, “Making data sharing work: the FCP/INDI experience,” *Neuroimage*, 2012.
- [183] P. T. Fox, and J. L. Lancaster, “Mapping context and content: the BrainMap model,” *Nature Reviews Neuroscience*, vol. 3, no. 4, pp. 319-321, 2002.
- [184] A. R. Laird, J. J. Lancaster, and P. T. Fox, “Brainmap,” *Neuroinformatics*, vol. 3,

no. 1, pp. 65-77, 2005.

- 
- [185] T. Onitsuka, M. E. Shenton, D. F. Salisbury, C. C. Dickey, K. Kasai, S. K. Toner, M. Frumin, R. Kikinis, F. A. Jolesz, and R. W. McCarley, "Middle and inferior temporal gyrus gray matter volume abnormalities in chronic schizophrenia: an MRI study," *Am J Psychiatry*, vol. 161, no. 9, pp. 1603, 2004.
- [186] J. Schröder, F. Wenz, L. Schad, K. Baudendistel, and M. Knopp, "Sensorimotor cortex and supplementary motor area changes in schizophrenia: a study with functional magnetic resonance imaging," *British Journal of Psychiatry*, vol. 167, no. AOU, pp. 197-201, 1995.
- [187] C. Exner, G. Weniger, C. Schmidt-Samoa, and E. Irle, "Reduced size of the pre-supplementary motor cortex and impaired motor sequence learning in first-episode schizophrenia," *Schizophrenia research*, vol. 84, no. 2, pp. 386-396, 2006.
- [188] G. R. Kuperberg, T. Deckersbach, D. J. Holt, D. Goff, and W. C. West, "Increased temporal and prefrontal activity in response to semantic associations in schizophrenia," *Archives of General Psychiatry*, vol. 64, no. 2, pp. 138, 2007.
- [189] P. R. Hof, V. Haroutunian, V. L. Friedrich Jr, W. Byne, C. Buitron, D. P. Perl, and K. L. Davis, "Loss and altered spatial distribution of oligodendrocytes in the superior frontal gyrus in schizophrenia," *Biol Psychiatry*, vol. 53, no. 12, pp. 1075-1085, 2003.
- [190] B. Crespo-Facorro, J.-J. Kim, N. C. Andreasen, D. S. O'Leary, H. J. Bockholt, and V. Magnotta, "Insular cortex abnormalities in schizophrenia: a structural magnetic resonance imaging study of first-episode patients," *Schizophrenia research*, vol. 46, no. 1, pp. 35-43, 2000.
- [191] M. E. Shenton, C. C. Dickey, M. Frumin, and R. W. McCarley, "A review of MRI findings in schizophrenia," *Schizophrenia research*, vol. 49, no. 1-2, pp. 1, 2001.
- [192] L. Beltrachini, A. Blenkman, N. von Ellenrieder, A. Petroni, H. Urquina, F. Manes, A. Ibáñez, and C. Muravchik, "Impact of head models in N170 component source imaging: results in control subjects and ADHD patients." p. 012019.
- [193] L. Kooistra, J. J. van der Meere, J. D. Edwards, B. J. Kaplan, S. Crawford, and B. G. Goodyear, "Preliminary fMRI findings on the effects of event rate in adults with ADHD," *Journal of neural transmission*, vol. 117, no. 5, pp. 655-662, 2010.
- [194] S. Carmona, O. Vilarroya, A. Bielsa, V. Tremols, J. Soliva, M. Rovira, J. Tomas, C. Raheb, J. Gispert, and S. Batlle, "Global and regional gray matter reductions in ADHD: a voxel-based morphometric study," *Neuroscience letters*, vol. 389, no. 2, pp. 88, 2005.

- 
- [195] M. F. Schneider, C. M. Krick, W. Retz, G. Hengesch, P. Retz-Junginger, W. Reith, and M. Rösler, “Impairment of fronto-striatal and parietal cerebral networks correlates with attention deficit hyperactivity disorder (ADHD) psychopathology in adults—a functional magnetic resonance imaging (fMRI) study,” *Psychiatry research. Neuroimaging*, vol. 183, no. 1, pp. 75, 2010.
- [196] R. A. Barkley, “Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD,” *Psychological bulletin*, vol. 121, no. 1, pp. 65, 1997.
- [197] E. G. Willcutt, A. E. Doyle, J. T. Nigg, S. V. Faraone, and B. F. Pennington, “Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review,” *Biol Psychiatry*, vol. 57, no. 11, pp. 1336-1346, 2005.