


6-4-2015

An Integrated Neuroimaging Approach for the Prediction and Analysis of Alzheimer's Disease and its Prodromal Stages

Qi Zhou
qzhou003@fiu.edu

DOI: 10.25148/etd.FIDC000070

Follow this and additional works at: <https://digitalcommons.fiu.edu/etd>

 Part of the [Electrical and Computer Engineering Commons](#), and the [Neuroscience and Neurobiology Commons](#)

Recommended Citation

Zhou, Qi, "An Integrated Neuroimaging Approach for the Prediction and Analysis of Alzheimer's Disease and its Prodromal Stages" (2015). *FIU Electronic Theses and Dissertations*. 2229.
<https://digitalcommons.fiu.edu/etd/2229>

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

AN INTEGRATED NEUROIMAGING APPROACH FOR THE PREDICTION AND
ANALYSIS OF ALZHEIMER'S DISEASE AND ITS PRODROMAL STAGES

A dissertation submitted in partial fulfillment of the
requirements for the degree of
DOCTOR OF PHILOSOPHY
in
ELECTRICAL ENGINEERING
by
Qi Zhou

2015

To: Dean Amir Mirmiran
College of Engineering and Computing

This dissertation, written by Qi Zhou, and entitled An Integrated Neuroimaging Approach for the Prediction and Analysis of Alzheimer's Disease and its Prodromal Stages, having been approved in respect to style and intellectual, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

Armando Barreto

Mercedes Cabrerizo

Ranjan Duara

Sakhrat Khizroev

Naphtali David Rishe

Malek Adjouadi, Major Professor

Date of Defense: June 4, 2015

The dissertation of Qi Zhou is approved.

Dean Amir Mirmiran
College of Engineering and Computing

Dean Lakshmi N. Reddi
University Graduate School

Florida International University, 2015

© Copyright 2015 by Qi Zhou

All rights reserved.

DEDICATION

I dedicate this dissertation to my parents and sister. Without their patience, understanding, support, and most of all love, the completion of this work would not have been possible.

ACKNOWLEDGMENTS

I would like to express my deepest appreciation and gratitude to my major professor, Dr. Malek Adjouadi, for his patience, guidance and for supporting me during these past four years. I also have to thank all my PhD committee members, Dr. Armando Barreto, Dr. Mercedes Cabrerizo, Dr. Ranjan Duara, Dr. Sakhrat Khizroev and Dr. Naphtali David Rishe for their helpful advice and suggestions. I would also like to specially thank Dr. Ranjan Duara, Warren Barker from the Wien Center for Alzheimer's disease and Memory Disorders at Mount Sinai Medical Center (Miami Beach) and Dr. David Loewenstein from the University of Miami for providing me with data and very insightful discussions and inputs to my research. I would also like to express my gratitude to all my teachers who instructed me and made me a better person.

I am highly indebted to my former colleagues Dr. Mohammed Goryawala and Dr. Jin Wang for sharing their experiences and research ideas with me. I am grateful to my colleagues at the Center for Advanced Technology and Education (CATE) for their kind advice and suggestions. I would also like to thank the CATE lab manager Ms. Niovi Rojas for the technical support she provided me.

Particularly, I have to thank my beloved family and friends for their love, unwavering supports and continuous encouragements throughout my graduate studies.

Finally, I am grateful for the support provided in the last year through the Dissertation Year Fellowship from the University Graduate School at Florida International University; the financial support provided was critical for completing my work. I would also like to recognize the support of the National Science Foundation under grants CNS-0959985, HRD-0833093, CNS-1042341, IIP-1338922 and IIP-

1230661. The philanthropic support provided by the Ware Foundation is also greatly appreciated. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012).

ABSTRACT OF THE DISSERTATION

AN INTEGRATED NEUROIMAGING APPROACH FOR THE PREDICTION AND
ANALYSIS OF ALZHEIMER'S DISEASE AND ITS PRODROMAL STAGES

by

Qi Zhou

Florida International University, 2015

Miami, Florida

Professor Malek Adjouadi, Major Professor

This dissertation proposes to combine magnetic resonance imaging (MRI), positron emission tomography (PET) and a neuropsychological test, Mini-Mental State Examination (MMSE), as input to a multidimensional space for the classification of Alzheimer's disease (AD) and its prodromal stages including amnesic MCI (aMCI) and non-amnesic MCI (naMCI). An assessment is provided on the effect of different MRI normalization techniques on the prediction of AD. Statistically significant variables selected for each combination model were used to construct the classification space using support vector machines. To combine MRI and PET, orthogonal partial least squares to latent structures is used as a multivariate analysis to discriminate between AD, early and late MCI (EMCI and LMCI) from cognitively normal (CN)s. In addition, this dissertation proposes a new effective mean indicator (EMI) method for distinguishing stages of AD from CN. EMI utilizes the mean of specific top-ranked measures, determined by incremental error analysis, to achieve optimal separation of AD and CN.

For AD vs. CN, the two most discriminative volumetric variables (right hippocampus and left inferior lateral ventricle), when combined with MMSE scores, provided an average accuracy of 92.4% (sensitivity: 84.0%; specificity: 96.1%). MMSE scores were found to improve classification accuracy by 8.2% and 12% for aMCI vs. CN and naMCI vs. CN, respectively. Brain atrophy was almost evenly seen on both sides of the brain for AD subjects, which was different from right side dominance for aMCI and left side dominance for naMCI. Findings suggest that subcortical volume need not be normalized, whereas cortical thickness should be normalized either by intracranial volume or the mean thickness. Furthermore, MRI and PET had comparable predictive power in separating AD from CN. For the EMCI prediction, cortical thickness was found to be the best predictor, even better than using all features together. Validation with an external test set demonstrated that best of feature-selected models for the LMCI group was able to classify 83% of the LMCI subjects. The EMI-based method achieved an accuracy of 92.7% using only MRI features. The performance of the EMI-based method along with its simplicity suggests great potential for its use in clinical trials.

TABLE OF CONTENTS

CHAPTER	PAGE
1. INTRODUCTION	1
1.1 Alzheimer's Disease	1
1.2 Biomarkers of Alzheimer's Disease	3
1.3 Research Purpose and Significance	5
1.4 Literature Survey	7
2. CONSTRUCTING AN OPTIMAL DECISIONAL SPACE	9
2.1 Background	9
2.2 Methodology	10
2.2.1 Subjects	10
2.2.2 MRI Protocol and Analysis	13
2.2.3 Feature Extraction and Variable Selection	14
2.2.4 SVM based Classification Experiments	17
2.3 Results	18
2.3.1 Rank of Variables	18
2.3.2 Best Variable Set based on Incremental Error Analysis	21
2.3.3 Reliability of Rank and Sample Size	25
2.3.4 Comparative Analysis	26
2.4 Discussion	28
3. SIGNIFICANCE OF NORMALIZATION ON ANATOMICAL MEASURES	32
3.1 Methodology	32
3.1.1 Subjects	32
3.1.2 Feature Extraction and Incremental Error Analysis	33
3.1.3 Normalization and Classification Experiment	34
3.2 Results and Discussions	35
3.2.1 Classification Performance and Model Selection	35
3.2.2 Univariate Analysis of Anatomical Measures	41
3.2.3 Spatial Distribution of Subjects Under the Best Model	49
3.2.4 Model Efficiency Estimation and Normalization	50
4. CLASSIFICATION AND PREDICTION BY PROPOSED EFFECTIVE MEAN INDICATOR	52
4.1 Background	52
4.2 Methodology	52
4.2.1 Inclusion and Diagnostic Criteria	52
4.2.2 MRI Acquisition and Analysis	54
4.2.3 Preprocessing of Data	54

4.2.4	EMI Calculation	55
4.2.5	Validation and Incremental Error Analysis.....	56
4.3	Results	58
4.3.1	Incremental error analysis	58
4.3.2	Histogram Plots, Boxplots and Scatter Plots of EMI.....	62
4.4	Discussion.....	68
4.4.1.	Flexibility in Balancing Between Performances.....	71
4.4.2.	Future Potential of EMI Method	74
5.	MULTIVARIATE ANALYSIS OF MRI AND PET (FDG AND 18F-AV-45).....	75
5.1	Background.....	75
5.2	Methodology.....	77
5.2.1	Dataset.....	77
5.2.2	Inclusion and diagnostic criteria	77
5.2.3	MRI and PET	78
5.2.4	Image analysis.....	80
5.2.5	Preprocessing of data	86
5.2.6	Multivariate data analysis.....	87
5.2.7	Validation with external test set.....	89
5.3	Results	90
5.3.1	Significance of features in separating groups	90
5.3.2	Predictive power of OPLS models	91
5.3.3	Model validation with external test set	93
5.4	Discussion.....	99
5.4.1	Significance of MRI and PET features	99
5.4.2	Model efficiency with OPLS.....	101
5.4.3	Model validation with external test set	104
6.	CONCLUSION AND FUTURE WORK	106
6.1	Conclusions from Analyzing Methods.....	106
6.1.1	Combining MRI with Neuropsychological Test (MMSE).....	106
6.1.2	Investigating Normalization Effect	107
6.1.3	Combining MRI with PET	108
6.1.4	EMI Approach.....	109
	REFERENCES	111
	VITA.....	127

LIST OF TABLES

TABLE	PAGE
Table 1: Participant demographics and characteristics in this study	12
Table 2: Significant variables for naMCI vs. CN and aMCI vs. CN	18
Table 3: Significant variables for AD vs. CN	19
Table 4: Summary of best performance cases based on incremental error analysis	24
Table 5: Summary of top rank variables with varying dataset size	26
Table 6: Performance comparison of different methods	27
Table 7: Demographic and neuropsychological characteristics of subjects	32
Table 8: Normalization Measures	34
Table 9: Classification performances on raw data	37
Table 10: Classification performances on normalized data	38
Table 11: Performance comparison of different methods	40
Table 12: Univariate analysis of subcortical volumes using different normalization approaches for AD vs. CN*	43
Table 13: Univariate analysis of surface area for left and right hemisphere*	44
Table 14: Univariate analysis of cortical thickness for left and right hemisphere*	46
Table 15: Demographics and characteristics of all participants	53
Table 16: Classification performance of the EMI-based method with the determined top 6, 31 and 3 features, as denoted by the maximal points shown in Figure 12 for AD vs. CN, LMCI vs. CN and EMCI vs. CN, respectively	62
Table 17: Descriptive statistics of the EMI	63
Table 18: Performance comparison of the proposed EMI-based method with other methods for predicting stages of Alzheimer's disease	73
Table 19: Demographics and characteristics of participants	79
Table 20: Statistical significance of cortical volumes by lobes	81

Table 21: Statistical significance of cortical thickness by lobes.....	82
Table 22: Statistical significance of surface area by lobes	83
Table 23: Statistical significance of subcortical volumes by lobes	84
Table 24: Statistical significance of hippocampus subfields and PET features.....	85
Table 25: Summary of predictive power (Q2Y) for all full models.....	93
Table 26: Classification performance (accuracy, sensitivity and specificity) for validation of best EMCI and LMCI models using external test set.....	99

LIST OF FIGURES

FIGURE	PAGE
Figure 1: General structure of the classification approach	11
Figure 2: Representation of all significant regions of interest (ROIs) (excluding morphometric variables and MMSE) using different colors for (a) AD vs. CN (b) aMCI vs. CN (c) naMCI vs. CN	20
Figure 3: Incremental error analysis performance of accuracy with standard deviation (S.D.) indicated as error bar (a) AD vs. CN (b) naMCI vs. CN and aMCI vs. CN with only one side S.D. shown to avoid overlapping. (For naMCI vs. CN, though only 3 variables were shown as significant, the 4th ranked variable was included to check if the accuracy continues to increase.)	23
Figure 4: Representation of the most significant regions that combines with MMSE scores constitute the optimal decisional space for a) AD vs. CN b) aMCI vs. CN c) naMCI vs. CN	25
Figure 5: A specific case displaying the distribution of the half subject population of the testing set in the context of the first three principle variables. Out of the 93 subjects shown and used for testing the SVM classifier only 5 misclassifications are seen which are shown by solid dots. This is a typical case of classification approach for AD vs. CN.	30
Figure 6: Representation of the top 5 significant subcortical volumes based on raw data in Table 12. (A) Superior view (B) Lateral view	47
Figure 7: Representation of the top 5 significant cortical thickness based on raw data in Table 14. (A) Left hemisphere (B) Right hemisphere	47
Figure 8: Representation of all significant surface area based on raw data in Table 13. (A) Left hemisphere (B) Right hemisphere	48
Figure 9: Representation of all significant surface area based on total-area normalized data on Table 13. (A) Left hemisphere (B) Right hemisphere.....	48
Figure 10: Representation of the whole dataset for the model of MMSE + subcortical volume, for a typical classification run under this model	49
Figure 11: General structure of the proposed EMI-based method.....	58
Figure 12: Incremental error analysis performance in terms of accuracy as a function of the number of top-ranked measures included for (A) AD vs. CN. (B) LMCI vs. CN. (C) EMCI vs. CN. The accuracy is the average obtained after 50	

independent experiments based on 10-fold cross validation with the maximal point highlighted with a big dot.	61
Figure 13: Histogram plots of the EMI using previously validated number of top-ranked measures for (top left) AD and CN. (top right) LMCI and CN. (bottom) EMCI and CN. In order to present the separation thresholds obtained from the training process of the 50 independent experiments, the maximum and minimum thresholds are plotted as blue lines.	65
Figure 14: Boxplots of the EMI using previously validated number of top-ranked measures for (left) AD and CN. (middle) LMCI and CN. (right) EMCI and CN. In order to present the separation thresholds obtained from the training process of the 50 independent experiments, the maximum and minimum thresholds are plotted as black dots.	66
Figure 15: Scatter plots of the EMI using previously validated number of measures against age for (A) AD vs. CN. (B) LMCI vs. CN. (C) EMCI vs. CN. In order to present the separation thresholds obtained from the training process of the 50 independent experiments, the maximum and minimum thresholds are plotted as blue lines.	68
Figure 16: Scatter plots of best full models as indicated in Table 25 for separation between (A) AD and CN. (B) LMCI and CN (C) EMCI and AD.	94
Figure 17: Predictive power (Q ² _Y) of feature-selected models with p representing feature-selected models based on p values, and VIP denoting feature-selected models based on VIP scores. All of the feature selection techniques were applied to all features except for EMCI_CT, which was applied only to cortical thickness as it was shown previously to have the highest predictive power of EMCI: (top) predictive power of p value-based and VIP-based feature-selected models of AD and LMCI; (bottom) Predictive power of p value-based and VIP-based feature-selected models of EMCI using all features or cortical thickness only.	95
Figure 18: Scatter plots of best full models validated with external test set. (A) Testing best EMCI model with external EMCI and CN subjects. (B) Testing best AD model with external EMCI and CN subjects. (C) Testing best LMCI model with external LMCI and CN subjects. (D) Testing best AD model with external LMCI and CN subjects.	97
Figure 19: Scatter plots of best feature-selected models validated with external test set. (A) Testing best feature-selected EMCI model with external EMCI and CN subjects. (B) Testing best feature-selected AD model with external EMCI and CN subjects. (C) Testing best feature-selected LMCI model with external LMCI and CN subjects. (D) Testing best feature-selected AD model with external LMCI data and CN subjects.	98

ABBREVIATIONS AND ACRONYMS

AD	Alzheimer's disease
ADNI	Alzheimer's Disease Neuroimaging Initiative
aMCI	Amnesic mild cognitive impairment
AV-45	18F-AV-45 (florbetapir)
APOE	Apolipoprotein E
CA	Cornu Ammonis
CDR	Clinical Dementia Rating
CN	Cognitive normal control
CSF	Cerebrospinal fluid
CT	Cortical thickness
CV	Cortical volume
DLB	Dementia with Lewy body
EEG	Electroencephalogram
EMCI	Early mild cognitive impairment
EMI	Effective mean indicator
FDG	Fluorodeoxyglucose
ICV	Intracranial volume
LMCI	Late mild cognitive impairment
MCI	Mild cognitive impairment
MMSE	Mini-mental state examination
MRI	Magnetic resonance imaging
naMCI	Non-amnesic mild cognitive impairment

OPLS	Orthogonal projections to latent structures
PCA	Principal component analysis
PET	Positron emission tomography
PLS	Partial least square
ROI	Regions of interest
SA	Surface area
SPECT	Single-photon emission computerized tomography
SUVR	Standardized uptake value ratio
SV	Subcortical volume
SVM	Support vector machine
UVS	Unit variance scaling
WMH	White matter hypointensities

CHAPTER 1

INTRODUCTION

1.1 Alzheimer's Disease

Nowadays more and more old people are suffering from Alzheimer's disease (AD) as a result of prolonged life expectancy (Duchesne, et al., 2008). AD is a progressively neurodegenerative disease, and is the most common form of dementia, accounting for 60% to 70% of the dementia cases (Organization, 2012). It is a syndrome that affects memory, thinking orientation, language, judgment, comprehension and so on. Estimates from the Alzheimer Association (alz.org) as of March 2012 indicate that 5.4 million Americans are diagnosed with AD, and over 95% of this population are 65 years of age or older. This estimate of AD population is expected to reach 16 million by 2050, a disturbing prospect. Also, nearly half of the population over 85 years of age is affected by AD (Organization, 2012).

AD patients display disease-related regional cerebral atrophy, which can be distinguished from normal aging (Fox and Schott, 2004; Ridha, et al., 2006). In AD, atrophy is often observed in regions which are closely related to neurodegeneration. Various studies have shown that atrophy in regions like the hippocampus (Apostolova, et al., 2012; Laakso, et al., 1995b; Leclerc and Abulrob, 2013; Scahill, et al., 2002), amygdala (Cuenod, et al., 1993; Laakso, et al., 1995b) and ventricles (Nestor, et al., 2008c; Thompson, et al., 2004b) is correlated to AD. Moreover, determination of the key atrophied regions across the entire

brain could be used as parameters for the delineation of AD patients from cognitively normal subjects (CN).

Transitional stage of AD is commonly referred as mild cognitive impairment (MCI), which is intermediate between cognitive normal (CN) controls and patients with AD. It is subdivided as early and late MCI (EMCI and LMCI) in the late two phases of Alzheimer's Disease Neuroimaging Initiative (ADNI) database (ADNI GO and ADNI 2). EMCI is characterized as very mild memory impairment that is intermediate between CN and LMCI (Herholz, et al., 2002; Landau, et al., 2012). Such division of MCI is preferred from the traditional classification of MCI as a general group as it makes differential diagnosis of AD more specific. It is expected that prediction of EMCI will be more challenging than for LMCI, while evidently it will be more beneficial to the patients if a correct diagnosis is made at the EMCI stage.

MCI were also divided into amnesic MCI (aMCI) and non-amnesic MCI (naMCI) and were studied separately as early findings suggest that these two different types of MCI have their own distinctive features that could delineate them (Ellison, et al., 2008; Nutter-Upham, et al., 2008). More specifically, this could help in predicting the conversion of MCI as has been reported by Ferman et al. that naMCIs were more likely to develop dementia with Lewy body (DLB) while patients with aMCI are more likely to convert to AD (Ferman, et al., 2013).

Mini-Mental State Examination (MMSE) is a neuropsychological test that is most often administered to screen patients for cognitive impairment and dementia (Folstein, et al., 1975). MMSE is used to judge the severity of cognitive impairment by administering 30 questions aimed at testing the subject's orientation to time and place, attention and calculation capabilities, as well as response to recall, language and complex commands. The frequent use of MMSE in clinical environments makes it interesting to investigate its discriminative power in classifying AD subjects as compared to MRI-based measures.

1.2 Biomarkers of Alzheimer's Disease

There are multiple biomarkers that have been shown effective in the diagnosis of AD and MCI. Magnetic resonance imaging (MRI) is commonly used for analyzing regional volumetric atrophy (Fennema-Notestine, et al., 2009; Fox, et al., 1999; Jack, et al., 1997; Wang, et al., 2011), cortical thinning (Dickerson, et al., 2009a; Eskildsen, et al., 2013), shape of specific regions (Achterberg, et al., 2014; Csernansky, et al., 2005; Wang, et al., 2007b) and structural/functional connectivity (Gour, et al., 2013; Haase, et al., 2013; Wang, et al., 2007a). In addition, extracellular beta-amyloid ($A\beta$) deposition (Hardy and Allsop, 1991b) and tau protein abnormalities (Mudher and Lovestone, 2002a) were reported as being the fundamental causes of the disease, making biomarkers capable of characterizing these critical features essential in the diagnosis of AD. Among these, Positron emission tomography (PET) imaging modality has been shown to be an efficient biomarker for evaluating amyloid plaque level and hypometabolism of the brain, which contain discriminative information for AD staging (Herholz, et al., 2002; Mosconi, et al.,

2008; Nordberg, et al., 2010a; Nordberg, et al., 2010b). Another biomarker is single-photon emission computerized tomography (SPECT), which measures regional cerebral perfusion (Hirao, et al., 2005; Jagust, et al., 2001; Johnson, et al., 1998).

Others include Cerebrospinal fluid (CSF) biomarker by analyzing level of β amyloid₁₋₄₂ (A β 42) and total tau protein (T-tau) of CSF (Blennow and Hampel, 2003; Hansson, et al., 2006; Mattsson, et al., 2009), EEG biomarker for estimating the power spectrum (Besthorn, et al., 1994; Locatelli, et al., 1998) and Apolipoprotein E (APOE) genotype indicating higher risk for AD with the presence of the ϵ 4 allele (Bickeböllner, et al., 1997; Ganzer, et al., 2003). These biomarkers have been widely used to guide clinicians for planning a course of action.

More recently, some studies have combined two or more of these biomarkers to explore their complementary information for improved discriminative power (Fan, et al., 2008b; Walhovd, et al., 2010a; Westman, et al., 2012a; Zhang, et al., 2011). For example, Zhang et al. combined biomarkers of MRI, PET and CSF using a proposed multimodal data fusion technique (Zhang, et al., 2011), and Westman et al. combined MRI and CSF using orthogonal partial least squares to latent structures (OPLS) analysis method (Westman, et al., 2012a). Specifically, some studies combined neuropsychological tests with medical imaging modalities. The cost-benefit of combining multiple biomarkers is of great concern as the use of multiple biomarkers often result in unaffordable costs (Westman, et al., 2012a).

1.3 Research Purpose and Significance

This dissertation investigates the classification of AD, aMCI and naMCI by combining subcortical volumes of MRI with a neuropsychological test (Mini-Mental State Examination (MMSE)), which is most often administered to screen patients for cognitive impairment and dementia. This study demonstrates the merits of MMSE and extends its use to the discrimination of different stages of AD when used in conjunction with select volumetric variables. To the best of our knowledge, this study is the first that investigates the impact of combining MRI at baseline with MMSE for the detection of AD, aMCI and naMCI using support vector machine (SVM) methodology. However, since MMSE as a cognitive score in combination with other biomarkers can introduce an unfair bias in the analysis, the work in this dissertation reflects the results under the two conditions which are with and without inclusion of MMSE. Another important contribution of this study is the development of a fully automated feature extraction technique, which in its initial step associates equal weights to each of the measured volumes, and yet as its outcome is a ranking of the volumes that can be used as variables in a multidimensional decisional space for optimal classification (Chapter 2).

In this dissertation, single-measure models and hierarchical models with and without normalization are both examined to find the optimal model. Single measure models include one of the regional MRI measures (subcortical volume, cortical thickness and surface area) or the neuropsychological test, Mini-Mental State Examination (MMSE). A hierarchical model combines two or more of the single-measure models to examine if the

interaction augments the classification process. The specific aims of this study are thus to determine: (1) the impact of neuropsychological test (MMSE) towards the classification; (2) the combination of regional measures and MMSE that yields the best classification performance; and (3) which normalization scheme should be employed to achieve a better classification performance (Chapter 3).

To combine MRI and PET, we included all applicable FreeSurfer segmentation and parcellation of MRI scans (i.e. subcortical volumes, cortical volumes, cortical thickness, surface areas and hippocampus subfields) to explore full patterns of atrophy reflected by MRI since, as it has been reported, volumetric measures of a few pre-defined structures cannot capture the spatio-temporal pattern of structural abnormalities in their entirety (Fan, et al., 2008b). We also included extracted features from AV-45 PET and FDG PET to investigate complimentary predictive information of MRI and the two PET imaging modalities using OPLS. All features were inputted to OPLS for a multivariate analysis of EMCI, LMCI and AD against cognitively normal controls (CN). The aims were: 1) To explore the individual and combined discriminative power of MRI and PET features in separating EMCI, LMCI and AD from CN and determine the best predictive models of EMCI, LMCI and AD; 2) to find out if feature selection techniques enhance prediction; and 3) to investigate efficiency of the models by validating with an external dataset and explore how best AD models would predict EMCI and LCMI subjects (Chapter 4).

Also a completely new approach without projection is proposed using effective mean indicator (EMI). EMI is a value associated with each subject and generated by averaging

a specific number of the top statistically significant MRI measures divided by the corresponding mean of each measure among CN subjects. Prediction is just as simple as comparing the EMI of the testing subjects with the threshold found by the training process (Chapter 5).

1.4 Literature Survey

Recently, some studies have combined two or more of these biomarkers to explore their complementary information for improved discriminative power (Fan, et al., 2008b; Walhovd, et al., 2010a; Westman, et al., 2012a; Zhang, et al., 2011). For example, Zhang et al. combined biomarkers of MRI, PET and CSF using a proposed multimodal data fusion technique (Zhang, et al., 2011), Westman et al. combined MRI and CSF using orthogonal partial least squares to latent structures (OPLS) analysis method (Westman, et al., 2012a) and Walhovd et al. and Daoqiang et al. reported that combination of MRI, PET and CSF biomarkers yields the most suitable and complementary indicators for the diagnosis of AD (or MCI) (Walhovd, et al., 2010c; Zhang, et al., 2011). The cost-benefit of combining multiple biomarkers is of great concern as the use of multiple biomarkers often result in unaffordable costs (Westman, et al., 2012a).

Important tasks to be considered in AD classification studies include the choice of parameters, the way these parameters ought to be combined, and in determining the pre-processing techniques to be employed in order to enhance the prospects of classification. Two essential questions that need to be addressed for AD classification studies are: (1) which regional MRI measures produced by Freesurfer are statistically significant for

classification of AD subjects?; and (2) which normalization approach should be employed to minimize bias due to differences in head size and brain structure in order to enhance the classification performance?

Westman and his colleagues have investigated some aspects of the aforementioned issues using a supervised multivariate data analysis using the orthogonal projections to latent structures (OPLS) model (Westman, et al., 2013). OPLS is similar to principal component analysis (PCA) as they both are linear decomposition techniques and project the original data to the found latent variables. The approach of this study is an extension of a previous study (Qin, et al., 2013), which proposes to construct for each classification model an optimal decisional space using the most statistically significant variables. The number of dimensions in the classifier is determined by an incremental error analysis, which in turn defines and ranks variables on their statistical significance to be used as input to an SVM-based classification process.

CHAPTER 2

CONSTRUCTING AN OPTIMAL DECISIONAL SPACE

2.1 Background

In literature, researchers have combined neuropsychological tests with medical imaging modalities. Among them, Ewers et al. combined the primary MRI and CSF biomarkers with neuropsychological tests to predict conversion from MCI to AD (Ewers, et al., 2012). With a population of 81 AD patients and 101 elderly control subjects, they determined that single-predictor models yielded comparable accuracies as multi-predictor models, with a prediction accuracy ranging from the mid-60s to a high of 68.5% when the entorhinal cortex is used as the single predictor. Gomar et al. investigated in a 2-year longitudinal study the usefulness of combining different variables drawn from a series of biomarkers with the inclusion of cognitive markers and other risk factors to likewise predict conversion from MCI to AD (Gomar, et al., 2011). Their study involved the use of brain volumes, CFS, along with other cognitive markers. Their findings suggest that cognitive markers at baseline are better suited as predictors for the conversion than the temporal neurobiological markers. Furthermore, they also suggest that sharp decline in functional ability is a better predictor for the conversion than the biomarkers. These findings add credence to the results obtained in this study, in that with the inclusion of neuropsychological data, accuracy in delineating AD from controls is shown to increase to over 90%. It is noted that in both of these studies, which focus more on the conversion

from MCI to AD, the volumetric measures of the different brain regions were selected manually, and both studies relied on the ADNI (Alzheimer's Disease Neuroimaging Initiative) public database. As such, the proposed study which provides an automated approach at ranking the neurobiological variables will augment and complement such findings, as reported in both of these studies, to reflect more globally patterns of structural and physiological abnormalities in their entirety (Fan, et al., 2008a), and with statistical context for a more meaningful choice of the different variables.

2.2 Methodology

The general structure of the proposed approach is presented in Figure 1, showing the main steps of the whole process from acquisition of the MRI scans, through the sorting and selection of variables or features that will constitute the decisional space for the classification process using the well-established SVM classifier. The proposed approach is also open to the use of other alternative classification algorithms such as artificial neural networks, optimal discriminant analysis, and so on. This study opted for SVM only for its implementation simplicity.

2.2.1 Subjects

A total of 309 participants were recruited from Wien Center for Alzheimer's Disease and Memory Disorders, Mount Sinai Medical Center, Miami Beach, FL as shown in Table 1 between 2005 and 2008. All participants have taken the Folstein Mini-Mental State Examination (Folstein, 1990) with a minimum score of 15. The study was approved by

the Mount Sinai Medical Center Institutional Review Board with informed consent provided by the subjects or legal representatives. All subjects had: (1) a neurological and medical evaluation by a physician; (2) Mini-Mental State Examination (MMSE) (3) a structural volumetrically acquired MRI scan of the brain. MMSE was used as the index of cognitive ability and sum of boxes from the Clinical Dementia Rating Scale (CDR-sb) was used clinically as the index of functional ability.

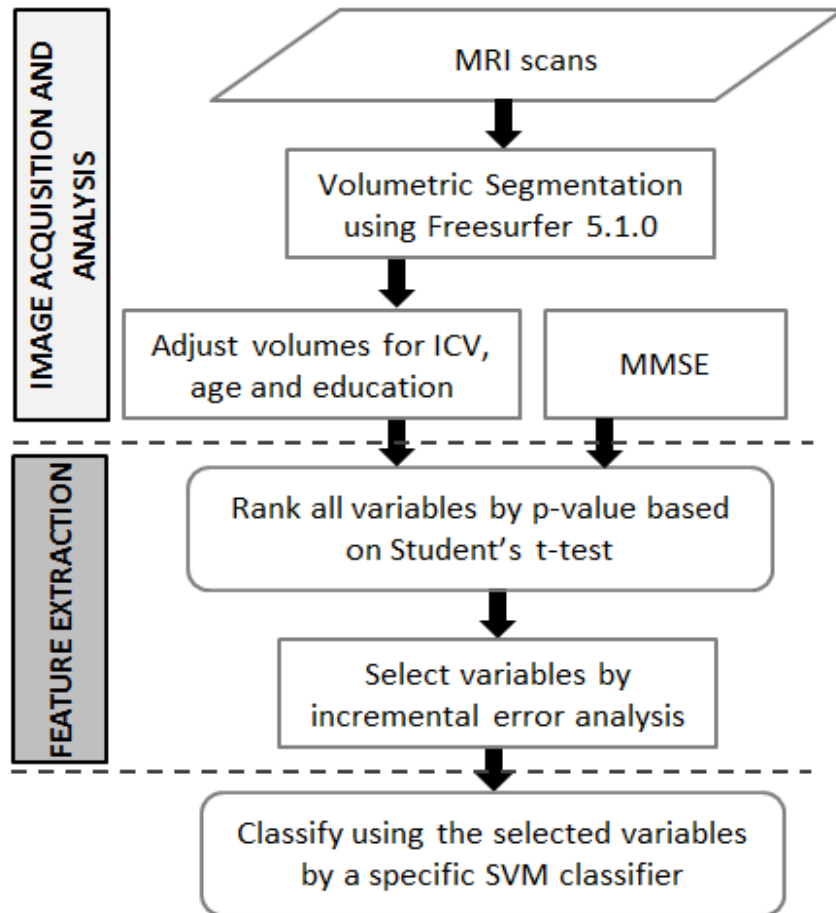


Figure 1: General structure of the classification approach

Table 1: Participant demographics and characteristics in this study

Characteristic	CN (n = 127)	naMCI (n = 56)	aMCI (n = 67)	AD (n = 59)	p-value*
Age	72.6 ± 6.0	74.1 ± 6.5	75.2 ± 6.8	79.8 ± 6.5	< 0.001
Gender(F/M)	91 / 36	36 / 20	36 / 31	33 / 26	ns [†]
Education (years)	14.4 ± 3.6	11.75 ± 3.6	13.3 ± 3.7	13.1 ± 3.9	0.035
MMSE	28.7 ± 1.4	26.9 ± 2.3	26.6 ± 2.5	22.7 ± 3.3	< 0.001

Unless otherwise noted, data are presented as mean ± S.D. *P-value based on Student's t-test between CN and AD unless otherwise specified and those less than significance level 0.05 are bolded. †A Fisher's exact test was performed between AD and CN, and p value shows that gender effect is not significant at significance level of 0.05.

The cognitive diagnosis was made using a combination of the physician's diagnosis and the neuropsychological diagnosis, as described previously (Loewenstein, et al., 2000). The etiological diagnosis was made by the examining physician. The diagnosis of cognitive normal (CN) required that the physician's diagnosis was CN and no cognitive test scores were ≥ 1.5 SD below age and education-corrected means. A probable AD diagnosis required a dementia syndrome and the National Institute of Neurological and Communicative Disorders and Stroke/Alzheimer's Disease and Related Disorders Association criteria for AD (McKhann, et al., 1984). The diagnosis of aMCI was rendered by 1) a clinical impression by the examining physician of a history of MCI but no significant functional impairment and did not meet Diagnostic and Statistical Manual of Mental Disorder-4th edition (DSM-IV) criteria (Association and DSM-IV., 1994) for dementia. This diagnosis was confirmed by a neuropsychological evaluation in which one or more tests of memory had to fall 1.5 SD or more below expected normative values.

The diagnosis of naMCI was rendered by 1) a clinical impression by the examining physician of a history of MCI but no significant functional impairment and did not meet DSM-IV criteria for dementia. This diagnosis was confirmed by a neuropsychological evaluation in which one or more tests of non-memory function (e.g., Trails B, Similarities, and Category Fluency) had to fall 1.5 SD or more below expected normative values but all tests of memory scored within normal limits.

2.2.2 MRI Protocol and Analysis

MRI scans were acquired on a 1.5-T machine (Siemen's Symphony, Iselin, N.J., USA, or General Electric, HDX, Milwaukee, Wisc., USA) using a proprietary 3D-magnetization-prepared rapid-acquisition gradient echo (3D MPRAGE) or 3D spoiled gradient echo sequences (FSPGR). Specifications for 3D MPRAGE include coronal sections with a 1.5 mm gap in thickness; section interval, 0.75 mm; TR, 2190 ms; TE, 4.38 ms; TI, 1100 ms; FA, 15°; NEX, 1; matrix, 256 × 256; FOV, 260 mm; bandwidth, 130 Hz/pixel; acquisition time, 9 minutes; phase-encoding direction, right to left. Specifications for 3D FSPGR were the following: 140 contiguous coronal sections of 1.2 mm thickness; contiguous images with no section interval; TR, 7.8 ms; TE, 3.0 ms; inversion recovery preparation time, 450 ms; flip angle, 12°; NEX, 1; matrix, 256 × 256; FOV, 240 mm; bandwidth, 31.25 Hz/pixel; acquisition time, 6–7 minutes; phase-encoding direction, right to left.

FreeSurfer pipeline (version 5.1.0) was applied to the MRI scans to produce 55 volumetric variables, including 45 subcortical regions (e.g. left lateral ventricle, corpus callosum anterior, right hippocampus, etc.) and 10 morphometric statistics (e.g. left hemisphere gray matter volume, total cortical volume, etc.). Out of the 45 volumetric variables, 4 of them, namely left white-matter-hypointensities (WMH), right WMH, left non-WMH, and right non-WMH were excluded since they were all characterized by zero values. Therefore, each MRI scan includes 41 regional and 10 morphometric volumes. It was determined that MRI scans from the two scanner machines did not change the variance of volume difference when comparing subcortical volumes (FreeSurfer segmentation) from the test-retest scans acquired in a fixed machine (Han, et al., 2006; Jovicich, et al., 2009), thus no correction is needed for scanner difference.

2.2.3 Feature Extraction and Variable Selection

AD patients suffer from cerebral atrophy, which can be distinguished from normal aging (Fox and Schott, 2004), and specific regions are more atrophied along the progression of AD. For example, studies have shown that hippocampal atrophy is more significant as disease progresses (Scahill, et al., 2002). Determination of the key atrophied/enlarged regions of interest (ROIs) that define the pattern of atrophy may help delineate AD and MCI from normal controls. To this end, a rigorous blind feature selection technique is proposed, where no prior assumptions of ROIs are assumed and with equal weights assigned to each of the volumetric measures so as to eliminate any bias. The outcome of this selection process is the determination of those statistically significant features that

form an optimal decisional space bound to yield accurate classification results of AD, MCI (aMCI, naMCI) and controls.

Freesurfer is a popular highly automated MRI image processing software widely used to generate regional measures from MRI scans. The advantages of Freesurfer over traditional manual segmentations and measures are its high automation and independence from operator subjectivity. Freesurfer is also accurate, precise and has been tested on large cohorts of studies in AD classification research (Cuingnet, et al., 2011; Ewers, et al., 2012; Qin, et al., 2013; Westman, et al., 2013).

Freesurfer pipeline version 5.1.0, widely used in AD research (Cuingnet, et al., 2011; Ewers, et al., 2012; Qin, et al., 2013; Westman, et al., 2013; Zhang, et al., 2011), was applied to all the MRI scans to produce 55 volumetric variables, including 45 volumetric measures of subcortical parcellation and 10 morphometric statistics. For cortical thickness, 34 regional variables were determined for each hemisphere, resulting in 68 variables for cortical thickness measures. Also, surface area was estimated from 35 regions of the brain for each hemisphere resulting in 70 measures for the entire brain.

All volumetric variables, but for intracranial (ICV), were adjusted for ICV, age and education as per Equation (2.1), as they were found to be significant factors as demonstrated in Table 1.

$$V_a = V_{ua} - G_{ICV} \cdot (V_{sICV} - V_{mICV}) - G_{EDU} \cdot (E_s - E_m) - G_{AGE} \cdot (A_s - A_m) \quad (2.1)$$

where V_a is the adjusted volume, V_{ua} is the unadjusted volume, V_{sICV} , E_s and A_s are the subject ICV, years of education and age (years), respectively; V_{mICV} , E_m and A_m are the corresponding means for all the control subjects. The gradients G_{ICV} , G_{EDU} and G_{AGE} were derived by a region specific regression against subject ICV, years of education and age of all the participants so that the regression is fully blinded to the classifications. As per Chiang et al. (Chiang, et al., 2011), the above regression also has the advantage that the regressing order of the 3 factors doesn't affect the results.

The adjusted volumes and ICV of the 51 volumetric variables are then combined with the MMSE score to generate a 52-variable vector discriminator for each subject. A Student's t-test is carried out on each of the 52 variables between AD (or MCI) and CN to determine the significance of each variable in the classification outcome and only those with a p-value lower than significance level (α) of 0.05 are selected and ranked.

It should also be noted that, even though atrophy is what is generally sought, statistical testing in this study considers both cases of atrophy and enlargement of brain regions, since volumetric enlargement (i.e. ventricles filled with cerebrospinal fluid) is also shown to be an important predictor of AD (Thompson, et al., 2004b).

Rank of the statistically significant variables provides an overall view of the discriminative power of each variable for each classification type. Selection of these optimal variables can be viewed as a dimensionality reduction problem, which is performed using an incremental error analysis. The result of this analysis is the

determination of how many of these top-ranked variables ought to be included in the classifier to yield the best classification performance. The proposed error analysis, which begins with the initial phase of using only the first top-ranked variable, is employed whereby an additional next-ranked variable is introduced in the SVM classifier at each subsequent phase and the corresponding classifier statistics (accuracy, sensitivity, specificity) are recorded each time.

2.2.4 SVM based Classification Experiments

SVM classifier is shown to be an effective classification tool of AD and MCI (Chaves, et al., 2009b; Kloppel, et al., 2008; Lopez, et al., 2009). The SVM as implemented in this study uses a sequential minimal optimization (SMO) scheme to implement a L1 soft-margin SVM classifier. SVM maps the original features via a kernel function to construct a maximum margin classifier in a high dimensional feature space. The kernel function used in this study is the Gaussian Radial Basis Function kernel (rbf) with an empirical scaling factor of 3.

In this study, all experiments were based on 2-fold cross validation, meaning half of the subjects data are used as training set and the other half as testing set. The training and testing sets were randomly assigned while the number of subjects with AD (or MCI) and CN in each set remain fixed. To limit the potential data partitioning error introduced by random data assignment and cross validation, the same experiment with random data assignment was run 50 times and the average performance was recorded.

2.3 Results

2.3.1 Rank of Variables

Based on the methodology described in section II (D), a Student's t-test was performed for AD vs.CN, aMCI vs. CN and naMCI vs. CN using the ICV-, education- and age-adjusted 52-variable vector discriminator. The variables found at significance level of 0.05 for MCI vs. CN and AD vs. CN were ranked and are given in Table 2 and Table 3, respectively, which show the importance of each variable in potentially discriminating different stages of AD. The anatomical distribution of these ROIs is shown in Figure 2.

Table 2: Significant variables for naMCI vs. CN and aMCI vs. CN

	Rank	Variables	p-value
naMCI	1	MMSE	< 0.001
	2	Right-Accumbens-area	0.013
	3	Right Inferior Lateral Ventricle	0.039
aMCI	1	MMSE	< 0.001
	2	Right-Hippocampus	< 0.001
	3	Left-Hippocampus	< 0.001
	4	Left-Amygdala	0.002
	5	Left Inferior Lateral Ventricle	0.023

From Table 2 and Table 3, it can be observed that MMSE as a cognitive measure ranks first for all classification types. Moreover, in comparison to MCI, the AD group is shown to have many more significant variables than the naMCI and aMCI groups, indicating atrophy is more serious and widely spread in the AD group. Table 2 also shows that hippocampus atrophy is dominant in aMCI subjects as both left and right hippocampus

regions rank as the second and third only after MMSE. This finding is consistent with the fact that one main characteristic of aMCI subjects is memory problem, which the hippocampus is found to be responsible for (Juottonen, et al., 1999). Whereas hippocampus region doesn't show as significant in naMCI group, Accumbens and ventricle regions do instead. The distinctive symptom of aMCI and naMCI (memory) is consistent with the regional atrophy in these preliminary results.

Table 3: Significant variables for AD vs. CN

Rank	Variables	p-value	Rank	Variables	p-value
1	MMSE	< 0.001	15	Right Lateral Ventricle	0.002
2	Right Hippocampus	< 0.001	16	Left Lateral Ventricle	0.002
3	Left Inferior Lateral Ventricle	< 0.001	17	Right Thalamus Proper	0.004
4	Left Amygdala	< 0.001	18	Right Ventral Diencephalon	0.007
5	Left Hippocampus	< 0.001	19	Left Thalamus Proper	0.007
6	Right Inferior Lateral Ventricle	< 0.001	20	Left Putamen	0.007
7	Right Hemisphere Cortex Volume	< 0.001	21	Corpus Callosum Central	0.010
8	Total Gray Volume	< 0.001	22	Right Accumbens Area	0.010
9	Right Amygdala	< 0.001	23	Corpus Callosum Posterior	0.011
10	Cortex Volume	< 0.001	24	Corpus Callosum Middle Anterior	0.020
11	Left Hemisphere Cortex Volume	< 0.001	25	Left Pallidum	0.026
12	Left choroid plexus	< 0.001	26	Corpus Callosum Anterior	0.028
13	Right choroid plexus	< 0.001	27	Subcortical Gray Volume	0.032
14	3 rd Ventricle	< 0.001	28	Corpus Callosum Middle Posterior	0.046

Anatomical visualization of all the significant ROIs (excluding morphometric variables) as shown in Figure 2 provides a general view of how the ROIs are distributed within the brain. The figure shows that atrophy for the AD group is widely seen for a number of regions on both hemispheres, while for aMCI atrophy is more dominant on the right side and for naMCI all significant atrophy is seen on the left side. This finding implies the difference of atrophy in terms of its extension and hemisphere dominance at different stages of AD.

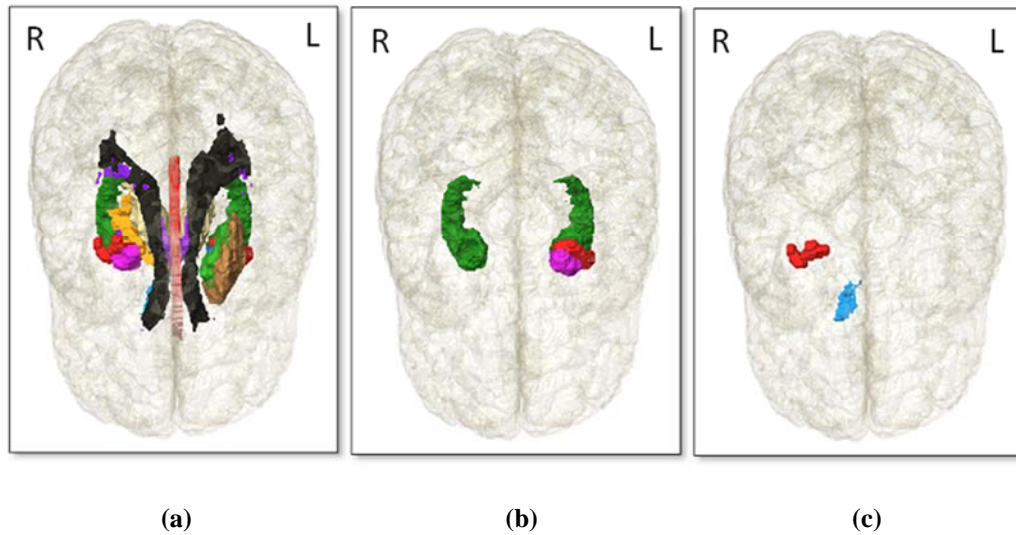


Figure 2: Representation of all significant regions of interest (ROIs) (excluding morphometric variables and MMSE) using different colors for (a) AD vs. CN (b) aMCI vs. CN (c) naMCI vs. CN

A closer inspection of the results shows that the top-ranked significant volumetric variables, e.g. hippocampus (de Leon, et al., 1989b; Du, et al., 2001; Pennanen, et al., 2004), ventricular (Nestor, et al., 2008b; Thompson, et al., 2004b), cortical (Du, et al., 2001; Juottonen, et al., 1999; Pennanen, et al., 2004) and amygdala (Cuenod, et al., 1993; Nestor, et al., 2008b), are all regions that have been proven to be effective predictors of

AD and/or MCI by other research groups. The convergence of these findings comes in support of the merits and usability of the ranking system developed in this study.

2.3.2 Best Variable Set based on Incremental Error Analysis

The incremental error analysis aims to determine how many top-ranked variables should be included in order to produce the best classification results. The classification of AD, aMCI and naMCI was performed starting from the first ranked variable, and incrementally adding the next best ranked variable until all significant variables were considered. The accuracy with error bars depicting the standard deviation of the 50 repetition runs is plotted in Figure 3, with the horizontal axis indicating at each step the number of top-ranked variables that were considered.

The classification accuracy curve based on the proposed incremental error analysis, displays a trend on how accuracy changes as more variables (one at a time) are included. For AD vs. CN, as shown in Figure 3a, the increasing trend of the curve stops when the first 3 variables (MMSE score, right hippocampus and left inferior lateral ventricle) are included, yielding a peak accuracy of 92.4%. As can be seen in Figure 3b, the accuracy peaks for the classification of naMCI when the top 2 ranked variables (MMSE and right hippocampus) are included with a peak accuracy of 74.9%; as for aMCI, the upward trend in accuracy is sustained up to the 3 significant variables (MMSE, right Accumbens area and right inferior lateral ventricle) reaching a peak accuracy of 74.1%.

The features that make up the optimal decisional space and hence produce the best classification results are summarized in Table 4, with and without the inclusion of the MMSE score. It can be observed that with MMSE, the performance improved significantly for all classification types with an accuracy increment of 14.2% for AD vs. CN, 8.2% for aMCI vs. CN and 12.0% for naMCI vs. CN. Such outcomes support the fact that the cognitive measure as determined from MMSE test is rather important in this type of classification. This difference in the accuracy could not otherwise be overcome with the inclusion of additional MRI-based volumetric measures.

To have an anatomical view of these optimal decisional spaces, the brain regions associated with the top-ranked volumetric features for each classification pair are plotted in Figure 4, which suggests that different brain regions could be linked to the different stages of AD.

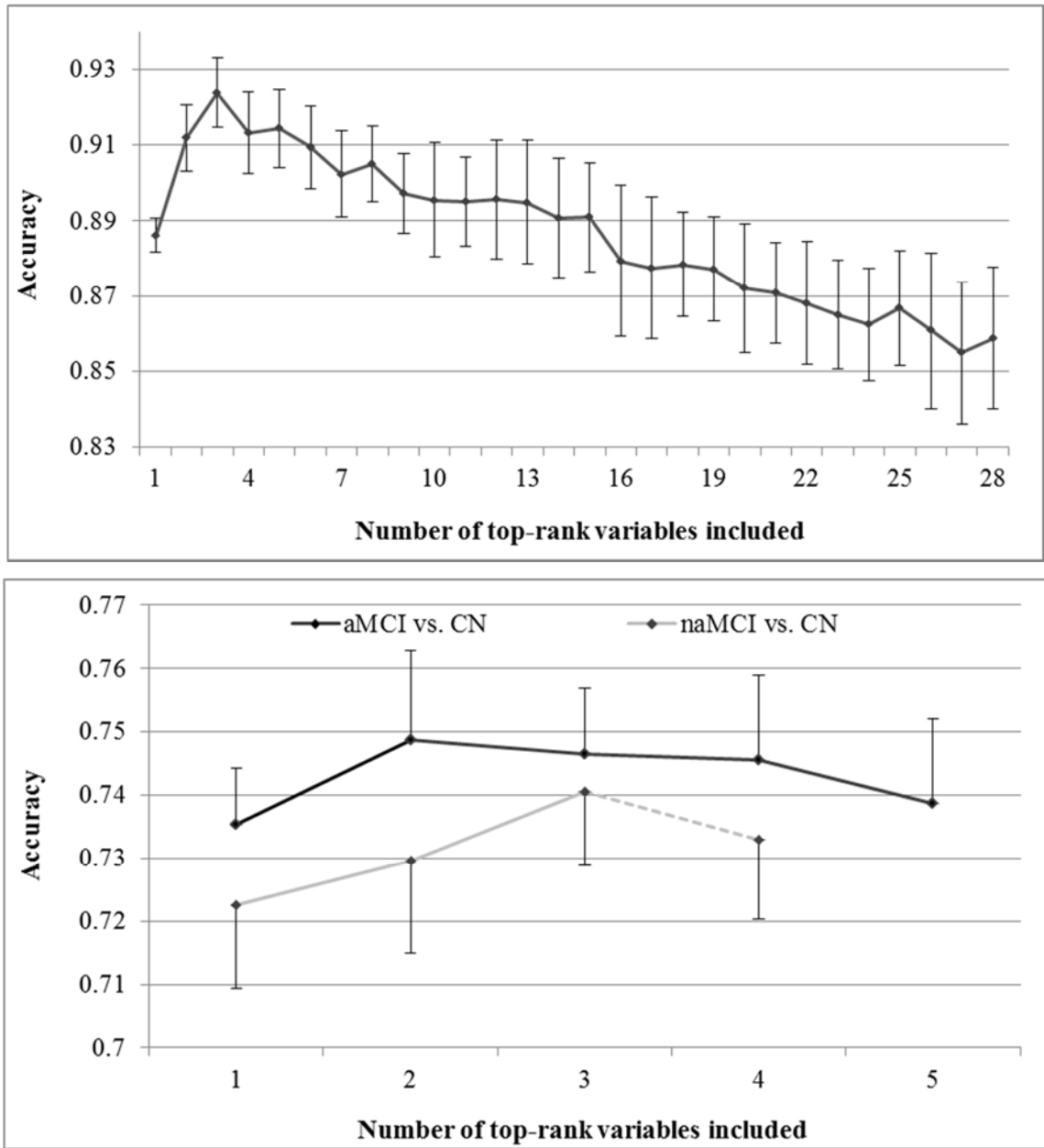


Figure 3: Incremental error analysis performance of accuracy with standard deviation (S.D.) indicated as error bar (a) AD vs. CN (b) naMCI vs. CN and aMCI vs. CN with only one side S.D. shown to avoid overlapping. (For naMCI vs. CN, though only 3 variables were shown as significant, the 4th ranked variable was included to check if the accuracy continues to increase.)

Table 4: Summary of best performance cases based on incremental error analysis

Classification Type	AD vs. CN		aMCI vs. CN		naMCI vs. CN	
	With MMSE	Without	With MMSE	Without	With MMSE	Without
Accuracy	92.4 (89.8–94.6)	78.2 (73.2–80.7)	74.9 (72.2–77.6)	66.7 (60.8–69.5)	74.1 (70.5–76.0)	62.1 (52.4–69.4)
Sensitivity	84.0 (77.8–88.2)	68.5 (60.9–74.4)	61.1 (53.3–68.7)	55.0 (49.2–62.5)	55.2 (50.0–62.5)	51.5 (41.1–62.5)
Specificity	96.1 (94.5–97.6)	82.7 (75.5–86.7)	83.4 (78.7–87.4)	72.5 (67.0–78.0)	82.3 (77.1–85.1)	70.3 (55.7–80.2)
Best Decisional Spaces	MMSE Right hippocampus Left inferior lateral ventricle		MMSE Right-Hippocampus		MMSE Right-Accumbens-area Right Inferior Lateral Ventricle	

2.3.3 Reliability of Rank and Sample Size

To evaluate the rank reliability of the variables which might be dependent on the dataset size, a test was performed by randomly reducing the dataset to 80% and to 50% of its original size. Then the ranks of each variable in 100 repetitions of the random experiment were summed and averaged to get the rank expectation. The final rank was determined by sorting the rank expectation of all variables from low to high and top 3 variables were recorded as shown in Table 5.

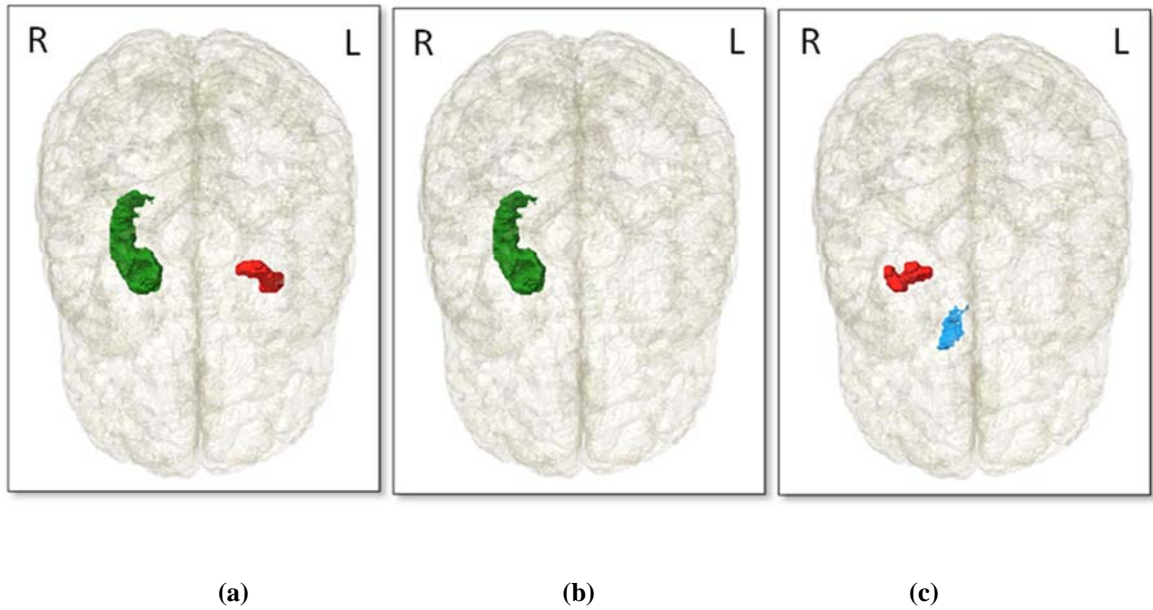


Figure 4: Representation of the most significant regions that combines with MMSE scores constitute the optimal decisional space for a) AD vs. CN b) aMCI vs. CN c) naMCI vs. CN

Table 5: Summary of top rank variables with varying dataset size

Dataset size	100%	80%	50%
AD vs. CN			
Rank of variables	MMSE	MMSE	MMSE
	RH	RH	RH
	LILV	LILV	LILV
aMCI vs. CN			
Rank of variables	MMSE	MMSE	MMSE
	RH	RH	RH
	LH	LH	LH
naMCI vs. CN			
Rank of variables	MMSE	MMSE	MMSE
	RAA	RAA	RAA
	RILV	RILV	RILV

RH = right hippocampus, LILV = left inferior lateral ventricle, LH = left hippocampus, RAA = right Accumbens area, RILV = right inferior lateral ventricle

2.3.4 Comparative Analysis

A comparison of classification performance with recent studies in literature is provided in Table 6, which also shows the detailed parameters of classification experiment for better comparison of the performance.

Table 6: Performance comparison of different methods

<i>Authors</i>	<i>Imaging Modality/ Biomarkers</i>	<i>Source of Data (AD/CN)</i>	<i>Repetition (cross validation)</i>	<i>Accuracy (%)</i>	<i>Sensitivity (%)</i>	<i>Specificity (%)</i>
Zhang et al., 2011(Zhang, et al., 2011)	MRI	ADNI(51/52)	10 (10 folds)	86.2	86.0	86.3
Zhang et al., 2011(Zhang, et al., 2011)	CSF	ADNI(51/52)	10 (10 folds)	82.1	81.9	82.3
Zhang et al., 2011(Zhang, et al., 2011)	PET	ADNI(51/52)	10 (10 folds)	86.5	86.3	86.6
Zhang et al., 2011(Zhang, et al., 2011)	MRI + PET + CSF	ADNI(51/52)	10 (10 folds)	93.2	93.0	93.3
Hinrichs et al., 2011(Hinrichs, et al., 2011)	MRI +PET	ADNI(48/66)	30 (10 folds)	87.6	78.9	93.8
Hinrichs et al., 2011(Hinrichs, et al., 2011)	MRI+PET+CSF+AP OE+Cognitive Scores	ADNI(48/66)	30 (10 folds)	92.4	86.7	96.6
Magnin et al., 2009(Magnin, et al., 2009a)	MRI	Private(16/22)	5000 (75% training / 25% testing)	94.5	91.5	96.6
Kloppel et al., 2008(Kloppel, et al., 2008)	MRI	(Group I) Private (20/20)	Leave-one-out	95.0	95.0	95.0
Kloppel et al., 2008(Kloppel, et al., 2008)	MRI	(Group II) Private (14/14)	Leave-one-out	92.9	100	85.7
Kloppel et al., 2008(Kloppel, et al., 2008)	MRI	(Group III) Private (33/57)	Leave-one-out	81.1	60.6	93.0
Walhovd et al., 2010(Walhovd, et al., 2010b)	MRI	ADNI(42/38)	N/A	82.5	81.6	83.3
Walhovd et al., 2010(Walhovd, et al., 2010b)	MRI + CSF	ADNI(42/38)	N/A	88.8	86.8	90.5
Cuingnet et al., 2011*(Cuingnet, et al., 2011)	MRI	ADNI(162/137)	-(2 folds)	-	81.0	95.0
Huang et al., 2000(Huang, et al., 2000)	EEG	Private (38/24)	-(-)	84.0	90.0	75.0
Proposed Study	MRI + MMSE	Private(127/59)	50 (2 folds)	92.4	84.0	96.1

*This paper by Cuingnet et al. compares ten methods and the best performance is shown here.

2.4 Discussion

The rank of the variables derived using the proposed feature selection process adds credence to the merits of combining structural MRI measures with the cognitive measure of MMSE for classifying AD, aMCI and non-aMCI from CN. Moreover, on the basis of the incremental error analysis, the top-ranked features as determined define the optimal decisional space on which group classifications (AD vs. CN, aMCI vs. CN and naMCI vs. CN) are carried out. It may be argued that the rank as derived in this study may vary as the size of the dataset under consideration is changed. A test trying to rerun the ranking of the variables by reducing the dataset size was done and recorded in Table 5, demonstrating that the rank doesn't change even when the dataset size is reduced by 50%. This indicates that the combination of variables found in this study is reliable, reproducible, and statistically meaningful even under a smaller subset of the data.

It can also be argued that incremental error analysis doesn't cover all the possible combinations of the significant variables as indicated in Table 2 and Table 3; thus there might be a combination that was not considered in the analysis that could have produced a better result. For AD vs. CN, there are 28 significant variables. All possible combinations are in the form of C_k^n (where n means the 28 significant variables, which could be combined 2 at a time, 3 at time etc. for different values of k, yielding different multidimensional decisional spaces). Such an exhaustive attempt at assessing all these combinations variables is not only unyielding, but rather unnecessary in light of the statistical meaningfulness which supports the proposed method of ranking these variables.

As can be observed from Figure 3, there were situations where performance increases slightly as other lower-ranked variables are considered. For example, when variable 8 is considered, the accuracy increases from 90.2% to 90.5% by 0.3% indicating that variable 8 does increase the classification performance and therefore should be considered as an added dimension in the decisional space. However, this could be due to the randomness of the subjects during the cross validation process using the SVM classifier. A slightly more favorable distribution of subjects can yield a relatively improved result. Consequently, the random distribution of the subject data was the primary reason for averaging the results of a large number of randomized runs (50) of the program for each experiment. On the other hand, the rank of the variable, especially lower-ranked variables, may suffer much variation from one dataset to another as they may intrinsically have very close mean differences identified by statistical test with close p values between the two groups under comparison. However this doesn't affect the meaningfulness of those very top-ranked variables. Furthermore, even though ranking variability may be observed, the top 3 ranked variables do not change. It should also be noted that determining the correct rank of the lower-ranked variables could be difficult and in some cases impractical due to the high unpredictability seen in these variables from one dataset to another.

The projection of subjects in the testing set on a decisional space based on the three aforementioned top-ranked variables for AD vs. CN is shown in Figure 5. A total of 5 misclassifications were observed which included 3 CN and 2 AD subjects. It can be seen that AD and CN groups are generally separable as they form two clearly distinct clusters,

especially the CN group, which is a denser cluster. This clustering outcome adds further credence to the importance of these three dimensions (variables).

Even though the use of MMSE score as a cognitive measure is limited in the sense that individuals may experience decline that is not captured, the preliminary results of this study suggests that MMSE improves classification results when combined with MRI measures. This implies that other neuropsychological tests such as verbal fluency test and delayed paragraph recall test may also contain complimentary information for improving classification of AD (or MCI) when used in conjunction with MRI measures.

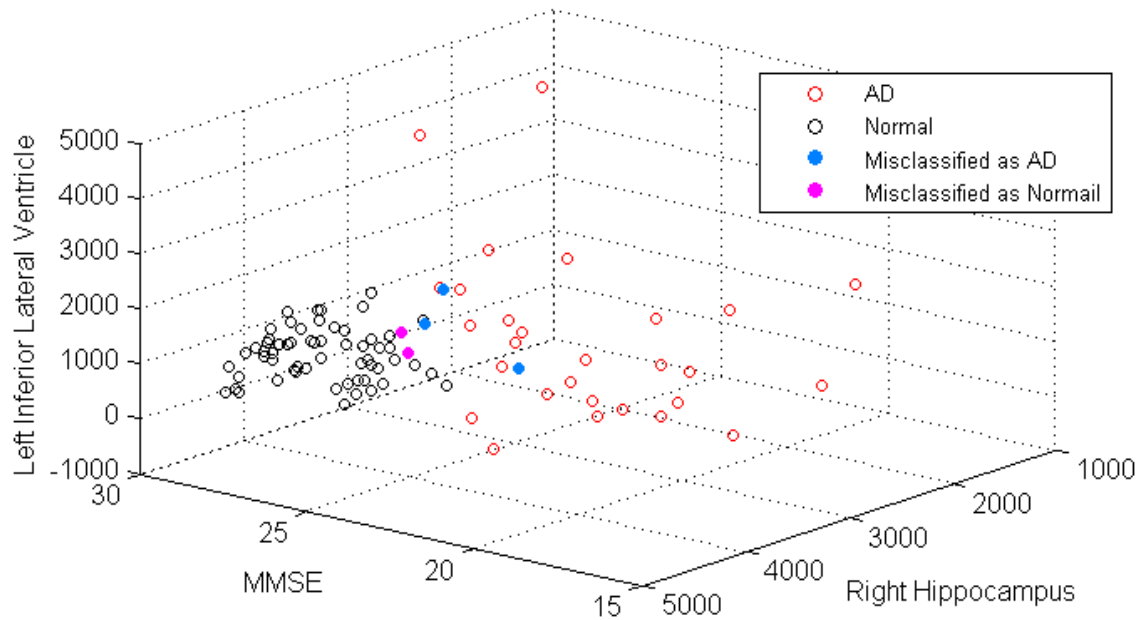


Figure 5: A specific case displaying the distribution of the half subject population of the testing set in the context of the first three principle variables. Out of the 93 subjects shown and used for testing the SVM classifier only 5 misclassifications are seen which are shown by solid dots. This is a typical case of classification approach for AD vs. CN.

From Table 6, a previous study using MMSE showed an accuracy of 92.3% for classifying AD from CN (Zhou, et al., 2014a). However, this study lacked the analysis for the MCI group. Table 6 indicates that the proposed technique using MMSE and MRI can yield competitive classification performance as those using two or more imaging modalities or biomarkers.

CHAPTER 3

SIGNIFICANCE OF NORMALIZATION ON ANATOMICAL MEASURES

3.1 Methodology

3.1.1 Subjects

A total of 189 subjects are included in this study as shown in Table 7. All participants are from the Wien Center for Alzheimer's Disease and Memory Disorders with the Mount Sinai Medical Center, Miami Beach, FL, USA. All subjects have taken the Folstein Mini-Mental State Examination (Folstein, et al., 1975) with a minimum score of 15 out of 30. The study was approved by the Mount Sinai Medical Center Institutional Review Board with informed consent provided by the subjects or legal representatives.

Table 7: Demographic and neuropsychological characteristics of subjects

	Age	Female/Male	MMSE
CN (n = 129)	72.9 ± 6.4	92 / 37	28.7 ± 1.4
AD (n = 60)	79.5 ± 6.9	34 / 26	22.6 ± 3.4
p	< 0.001	ns	< 0.001

Data Presented as mean ± S.D. where applicable. Two-way Student t test was used to test for age and MMSE and Fisher's exact test was used to test for gender.

All subjects had: (1) a neurological and medical evaluation by a physician; (2) a full battery of neuropsychological tests (Duara, et al., 2010), according to the National Alzheimer's Coordinating Center protocol, and the following additional tests: the Three-

Trial Fold Object Memory Evaluation (Fuld, 1982) and the Hopkins Verbal Learning Test ; as well as (3) a structural volumetrically acquired MRI scan of the brain. The Clinical Dementia Rating Scale (CDR-sb) was used as the index of functional ability, and the MMSE was used as the index of cognitive ability. The cognitive diagnosis was made using a combination of the physician's diagnosis and neuropsychological diagnosis. The etiological diagnosis was made by the examining physician. The diagnosis of CN required that the physician's diagnosis was CN and no cognitive test scores were ≥ 1.5 SD below age- and education-corrected means. A probable AD diagnosis required a dementia syndrome and the National Institute of Neurological Disorders and Stroke (NINDS)/Alzheimer's Association criteria for AD (McKhann, et al., 1984).

3.1.2 Feature Extraction and Incremental Error Analysis

All the variables in a given model are first ranked based on statistical significance between AD and CN. Following this ranking, an incremental error analysis is used whereby the SVM classifier is trained and tested adding a single variable at a time to the classifier to determine the combination of top-ranked variables that yield the optimal classification outcome. This rigorous blind feature selection technique differs from others as it does not rely on prior assumptions of regions of interest (ROI) and thus assigns equal weights to all the variables. The above process was performed on all models to compare their discriminative power and consequently identify the optimal model for AD classification. It should be noted that although regional atrophy among AD patients is what is generally sought, the statistical test considers both cases of atrophy and

enlargement of these specific brain regions, since volumetric enlargement can be experienced in regions like the ventricles, which has been shown to be important in differentiating AD and its prodromal stages (Apostolova, et al., 2012; de Leon, et al., 1989a; Thompson, et al., 2004b).

3.1.3 Normalization and Classification Experiment

To explore the effect of normalization on the classification performance, MRI measures are normalized by the widely accepted morphometric measures like intracranial volume (ICV) for regional subcortical volumes, ICV and mean cortical thickness of the subject for regional cortical thickness, and ICV and the total surface area of the subject for regional surface area. A summary of the normalization measures is presented in Table 8. ICV is derived from the MRI and is one among the 10 morphometric statistics obtained by the Freesurfer pipeline. Mean cortical thickness is estimated by averaging the thickness of all the 68 regions of the brain for each subject. Similarly, total surface area is the sum of all regional surface area measures for a given subject.

Table 8: Normalization Measures

MRI Measure	Morphometric Normalization Measure
Subcortical volumes (SV)	Intracranial Volume (ICV)
Cortical Thickness (CT)	Intracranial Volume (ICV) Mean Cortical Thickness
Surface Area (SA)	Intracranial Volume (ICV) Total Surface Area

Classification was performed using a support vector machine (SVM) classifier, which is shown to be effective as a classification tool for AD (Kloppel, et al., 2008; Lopez, et al., 2009; Magnin, et al., 2009a). The kernel function of the SVM used for this particular study is the Gaussian Radial Basis Function kernel (rbf) with a scaling factor (σ) of 3. All the classification results reported here are based on a 5-fold cross validation process. Each classification experiment was run 50 times, the results of which are averaged to evaluate the performance in terms of accuracy, sensitivity, specificity and precision.

3.2 Results and Discussions

3.2.1 Classification Performance and Model Selection

Single measure models using only one type of the regional measures or MMSE were created for subcortical volume, cortical thickness, surface area and neuropsychological data (MMSE) for both raw and normalized data. Hierarchical models were also created by combining two or more of the single models for both raw and normalized data. Feature selection based on statistical testing was performed for all the models created. The results of models with raw data are shown in Table 9 and the results for models with normalized data are shown in Table 10. All the results display an average of 50 runs with minimum and maximum values shown in parentheses.

Results of the different models are highly consistent as results of the 50 independent repetitions of classification fall within a small range as shown by the minimum and maximum values in Table 9 and Table 10. This small range is a clear indication of the

replicability of results, both essential attributes in any classification process. These results also indicate that MMSE is an important factor that should be included in the classification process. Inclusion of MMSE with other measures improves significantly the classification results. For example in the case of the optimal model, hierarchical model using subcortical volumes (SV) with the inclusion of MMSE resulted in an improvement of 9.2% as compared to using SV alone. In retrospect, an average improvement of 13.3% is seen on comparing analogous models with and without MMSE when using raw data and 12.8% when using normalized data.

The classification results given in Table 9 and Table 10 show that cortical thickness should be normalized by either the mean thickness of all the measured regions or ICV, while normalizing subcortical volumes to ICV doesn't have any significant effect. In a recent study, Westman et al. explored the normalization effect of regional MRI measures using orthogonal partial least square to latent structures (OPLS) models and concluded that both cortical thickness and subcortical volumes should not be normalized (Westman, et al., 2013). Both studies thus suggest that subcortical volumes should not be normalized to ICV. The divergence is seen in the normalization of cortical thickness. This could be potentially explained by the difference of the technique being used. Westman and his colleagues used an all variables inclusive model (OPLS) and the proposed method is feature selection based. The cause might be that normalization of cortical thickness brings down the variation of all the regions in general which OPLS model rely on but enhance variation in some specific regions that feature selection method might have selected. Thus, normalization of cortical thickness depends on the

processing technique used. Also, the divergence can be due to the subtle differences in the data that is used for the study.

Table 9: Classification performances on raw data

Model	Accuracy	Sensitivity	Specificity	Precision
MMSE	88.3 (87.3-89.4)	81.0 (76.7-81.7)	91.6 (91.5-94.6)	82.6 (81.3-87.6)
Subcortical volume (SV)	83.1 (81.5-85.2)	77.9 (75.0-80.0)	85.6 (83.0-88.4)	72.6 (69.2-77.6)
Cortical thickness (CT)	77.7 (76.2-78.9)	74.8 (73.3-76.7)	79.0 (77.4-80.7)	63.0 (59.9-68.0)
Surface area (SA)	71.4 (68.3-73.6)	58.7 (53.3-65.0)	77.2 (73.6-79.8)	55.0 (51.9-58.9)
Average	80.1	73.1	83.4	68.3

Hierarchical Model

MMSE + SV	92.3 (90.5-93.1)	88.2 (85.0-90.7)	94.2 (92.3-95.3)	88.3 (85.1-90.5)
MMSE + CT	91.4 (90.4-92.6)	85.3 (83.3-88.3)	94.2 (93.0-95.4)	87.8 (85.1-90.3)
MMSE + SA	88.6 (86.3-89.5)	76.3 (71.7-78.3)	94.3 (91.5-95.4)	87.1 (81.4-89.9)
CT + SV*	83.1 (81.5-85.2)	77.9 (75.0-80.0)	85.6 (83.0-88.4)	72.6 (69.2-77.6)
SA + CT + SV*	83.1 (81.5-85.2)	77.9 (75.0-80.0)	85.6 (83.0-88.4)	72.6 (69.2-77.6)
MMSE + SA + CT + SV**	92.3 (90.5-93.1)	88.2 (85.0-90.7)	94.2 (92.3-95.3)	88.3 (85.1-90.5)
Average	88.5	82.3	91.4	82.8

*The results of these models are the same as model of 'SV' since the variables extracted for the decisional space are the same as that for 'SV' model. ** This model gives identical results as the model of 'MMSE+SV' since variables extracted for the decisional space are the same as that for 'MMSE+SV'

Table 10: Classification performances on normalized data

Model	Accuracy	Sensitivity	Specificity	Precision
Subcortical volume (SV)	83.5 (82.0-84.7)	74.4 (71.7-76.7)	87.7 (95.3-90.0)	75.2 (72.0-79.0)
Cortical thickness (CT)	79.0 (77.8-80.4)	78.8 (75.0-81.7)	79.2 (77.5-80.6)	64.5 (61.8-87.4)
CT (Mean)*	78.9 (77.2-80.5)	78.4 (75.0-81.7)	79.2 (75.7-80.7)	64.6 (60.9-68.4)
Surface area (SA)	72.3 (68.8-75.2)	42.6 (35.0-48.3)	86.1 (82.1-89.2)	60.4 (50.8-65.3)
SA (Area)**	72.6 (70.3-75.1)	61.2 (58.3-63.3)	77.9 (75.1-81.4)	57.4 (52.9-61.8)
Average	77.3	67.1	82.02	64.4

Hierarchical Model

MMSE + SV	91.7 (90.0-93.1)	85.8 (81.7-88.3)	94.5 (93.1-95.4)	88.2 (85.0-90.7)
MMSE + CT	91.5 (89.4-93.2)	86.9 (81.7-90.0)	93.6 (90.8-96.1)	87.2 (82.7-90.9)
MMSE + CT (Mean)*	90.3 (89.2-91.1)	90.8 (89.6-91.7)	90.1 (88.2-91.7)	81.4 (78.5-83.9)
MMSE + SA	88.3 (87.3-88.9)	80.9 (76.7-81.7)	91.7 (91.4-93.8)	82.7 (81.1-86.1)
MMSE + SA (Area)**	88.6 (86.8-89.9)	80.9 (75.0-78.3)	94.2 (92.2-95.4)	86.9 (84.6-89.5)
CT + SV	83.1 (80.9-84.2)	75.8 (73.3-76.7)	86.5 (84.4-88.4)	73.3 (70.2-76.8)
CT + SA + SV	83.4 (81.0-85.7)	78.0 (75.0-80.0)	85.9 (83.0-89.1)	73.2 (68.3-69.4)
MMSE CT + SA + SV	91.7 (90.4-92.6)	86.0 (83.3-90.0)	94.4 (93.7-95.4)	88.4 (86.5-90.2)
Average	88.6	83.1	91.3	82.7

* Scaled by the mean thickness of the all the thickness measures

**Scaled by the total area of the all the measures

Since some models have very close performance in terms of the 4 recorded performance metrics (accuracy, sensitivity, specificity and precision), models that give more than 90% accuracy are considered as good models and are highlighted in gray in Table 9 and Table 10. Inclusion of additional measures does not guarantee a significant performance enhancement. A tradeoff exists between models with some displaying better accuracy at the cost of sensitivity and vice versa. In terms of accuracy, the model of ‘MMSE + SV’ is the best; whereas in terms of sensitivity, the model of “MMSE + CT (Mean)” is more appropriate.

A comparison of classification performance with recent studies in literature is provided in Table 11. The results indicate that the proposed technique using MMSE and MRI can yield competitive classification performance as those using two or more imaging modalities or biomarkers. As Westman and his team described the concept of cost-benefits to assess the increased cost of combining biomarkers as the potential limitation (Westman, et al., 2012b), the proposed approach has the advantage of low cost yet high accuracy. In addition, the results in this study are based on a larger cohort than most other studies in Table 11.

Table 11: Performance comparison of different methods

Authors	Imaging Modality/ Biomarkers	Source of Data (AD/CN)	Repetition (cross validation)	Accuracy (%)	Sensitivity (%)	Specificity (%)
(Zhang, et al., 2011)	MRI	ADNI(51/52)	10 (10 folds)	86.2	86	86.3
(Zhang, et al., 2011)	CSF	ADNI(51/52)	10 (10 folds)	82.1	81.9	82.3
(Zhang, et al., 2011)	PET	ADNI(51/52)	10 (10 folds)	86.5	86.3	86.6
(Zhang, et al., 2011)	MRI, PET, CSF	ADNI(51/52)	10 (10 folds)	93.2	93.0	93.3
(Hinrichs, et al., 2011)	MRI + PET	ADNI(48/66)	30 (10 folds)	87.6	78.9	93.8
(Hinrichs, et al., 2011)	MRI+PET+CSF+AP OE+Cognitive Scores	ADNI(48/66)	30 (10 folds)	92.4	86.7	96.6
(Magnin, et al., 2009a)	MRI	Private(16/22)	5000 (75% training / 25% testing)	94.5	91.5	96.6
(Kloppel, et al., 2008)	MRI	Group I (20/20)	Leave-one-out	95.0	95.0	95.0
(Kloppel, et al., 2008)	MRI	Group II (14/14)	Leave-one-out	92.9	100	85.7
(Kloppel, et al., 2008)	MRI	Group III (33/57)	Leave-one-out	81.1	60.6	93.0
(Walhovd, et al., 2010b)	MRI	ADNI(42/38)	N/A	82.5	81.6	83.3
(Walhovd, et al., 2010b)	MRI + CSF	ADNI(42/38)	N/A	88.8	86.8	90.5
*(Cuingnet, et al., 2011)	MRI	ADNI(162/137)	- (2 folds)	N/A	81.0	95.0
Proposed Study	MRI + MMSE	Private(129/60)	50 (5 folds)	92.3	88.2	94.2

*This paper by Cuingnet et al. compares ten methods and the best performance is shown here.

3.2.2 Univariate Analysis of Anatomical Measures

This section investigates how normalization affects the statistical significance of the variables that are used in the classification model. The effect of normalization can be determined by observing the change in the significance of the MRI measures when normalization is carried out. To illustrate the effect of normalization approaches on the statistical significance of region of interests (ROIs), univariate analysis was performed for subcortical volumes as shown in Table 12, and on surface area for left and right hemisphere respectively as shown in Table 13, and on cortical thickness for left and right hemisphere respectively as shown in Table 14. Univariate analysis was created for the two hemispheres separately for both cortical thickness and surface area in order to inspect the possible pattern differences between left and right hemisphere. In Table 12 – Table 14, the regions of the brain for which the significance of the variable differ between raw and normalized data are bolded. Please note that only those regions which show such a behavior for both the normalization techniques are highlighted in Table 13 and Table 14 .

Table 12 shows that ICV normalization to the subcortical volumes does not change the statistical significance of the variables, particularly for the top-ranked variables; suggesting that normalizing subcortical volumes with ICV might not be necessary, which is consistent with the conclusion made previously that subcortical volumes are not recommended to be normalized to ICV as seen from the results provided earlier in Table 13 and Table 14.

More importantly, subcortical volumes and cortical thickness show symmetry between the left and right hemispheres for the top-ranked variables as shown in Table 12 – Table 14. In other words, regions of the brain that are significant towards classification of AD subject are symmetrically located on either lobes of the brain. A typical example is seen in the top 5 ranked regions according to subcortical volumes which include both the right and left hippocampus and the right and left inferior lateral ventricles.

However, Table 13 shows that for the surface area there is almost no symmetry at all between the left and right hemispheres for both the raw and normalized data. This could possibly be explained by the fact that all variables found to be significant using surface area possess a p-value close to the significance level threshold (0.05). Another point to be noted is that for both raw and normalized data, surface area has a smaller number of significant variables and relatively high p-values, indicating that surface area may be generally regarded as a weaker biomarker of AD atrophy than the other two measures which are SV and CT.

Table 12: Univariate analysis of subcortical volumes using different normalization approaches for AD vs. CN*

Volumes	Raw		ICV	Volumes	Raw		ICV
	p values						
Right Hippocampus	< 0.00001	< 0.00001		Corpus Callosum Middle Anterior	< 0.0001	< 0.001	
Left Inferior Lateral Ventricle	< 0.00001	< 0.00001		Right Accumbens area	< 0.0001	< 0.01	
Left Hippocampus	< 0.00001	< 0.00001		Corpus Callosum Posterior	< 0.0001	< 0.01	
Left Amygdala	< 0.00001	< 0.00001		Right Thalamus Proper	< 0.0001	< 0.01	
Right Inferior Lateral Ventricle	< 0.00001	< 0.00001		Corpus Callosum Middle Posterior	< 0.0001	< 0.01	
Cortex Volume	< 0.00001	< 0.00001		White Matter Hypo-Intensities	< 0.0001	< 0.0001	
Left Hemisphere Cortex Volume	< 0.00001	< 0.00001		Left Accumbens area	< 0.001	< 0.001	
Right Hemisphere-Cortex Volume	< 0.00001	< 0.00001		Cerebral Spinal-Fluid(CSF)	< 0.001	< 0.00001	
Total Gray Volume	< 0.00001	< 0.00001		Right Ventral Diencephalon	< 0.001	ns	
3 rd Ventricle	< 0.00001	< 0.00001		Left Thalamus Proper	< 0.001	ns	
Right Amygdala	< 0.00001	< 0.00001		Non-White Matter Hypo-Intensities	< 0.001	< 0.01	
Right choroid plexus	< 0.00001	< 0.00001		Subcortical Gray Volume	< 0.01	< 0.05	
Right Lateral Ventricle	< 0.00001	< 0.00001		Optic Chiasm	< 0.01	< 0.01	
Left Lateral Ventricle	< 0.00001	< 0.00001		5 th Ventricle	< 0.05	< 0.05	

Table 13: Univariate analysis of surface area for left and right hemisphere*

Surface area	Left Hemisphere			Right Hemisphere		
Normalization	Raw	ICV	Total area	Raw	ICV	Total area
Bankssts	< 0.01	< 0.05	< 0.001	ns	ns	ns
Frontalpole	< 0.01	< 0.05	< 0.05	ns	ns	ns
Paracentral	< 0.05	< 0.01	< 0.01	ns	ns	ns
Transverse-temporal	ns	< 0.01	< 0.01	ns	ns	ns
Lingual	ns	ns	< 0.01	ns	ns	ns
Postcentral	ns	< 0.01	< 0.01	ns	ns	ns
Insula	ns	< 0.05	< 0.01	ns	ns	ns
Cuneus	ns	ns	< 0.05	ns	ns	< 0.05
Temporalpole	ns	ns	ns	< 0.01	< 0.01	< 0.001
Superior-frontal	ns	< 0.05	< 0.01	ns	< 0.05	< 0.01
Precentral	ns	ns	< 0.05	ns	< 0.05	< 0.01
Fusiform	ns	ns	< 0.05	ns	ns	< 0.01
Inferiortemporal	ns	ns	ns	ns	ns	< 0.01
Inferiorparietal	ns	ns	ns	ns	ns	< 0.05

*Two-way Student t test is used for univariate analysis with a significant level of 0.05 for p-value.

The regions of the brain which are determined to be statistically significant are displayed in Figure 6 through Figure 9. Figure 6 represents the top 5 significant subcortical volumes based on raw data. Figure 7 and Figure 8 represent the cortical regions of the brain which are found to be significant for AD classification using cortical thickness (CT) and surface area (SA) respectively on raw data. Figure 9 illustrates the change that is seen in the significant regions of the brain when surface area normalized to the total surface area is used as a measure, as compared to raw data as shown in Figure 8.

One interesting finding about cortical thickness in Figure 7 is that most of the significant regions belong to the temporal lobe, suggesting that the temporal lobe undergoes the most significant thickness change. This is consistent with the result found by some other studies (Dickerson, et al., 2009b; Rombouts, et al., 2000), particularly the finding that large degree of thinning of temporal cortical thickness seen in AD while thinning is relatively reserved in normal aging(Dickerson, et al., 2009b). The non-symmetric atrophy pattern of surface area can be easily observed anatomically in Figure 8 and Figure 9.

Table 14: Univariate analysis of cortical thickness for left and right hemisphere*

Cortical thickness Normalization	Left Hemisphere			Right Hemisphere		
	Raw	ICV	Mean CT	Raw	ICV	Mean CT
Superiortemporal	< 0.00001	< 0.00001	< 0.00001	< 0.00001	< 0.00001	< 0.00001
Entorhinal	< 0.00001	< 0.00001	< 0.00001	< 0.00001	< 0.00001	< 0.00001
Temporalepole	< 0.00001	< 0.00001	< 0.00001	< 0.00001	< 0.00001	< 0.00001
Inferiortemporal	< 0.00001	< 0.00001	< 0.01	< 0.00001	< 0.0001	< 0.01
Middletemporal	< 0.00001	< 0.00001	< 0.05	< 0.00001	< 0.0001	< 0.05
Parahippocampal	< 0.00001	< 0.00001	< 0.01	< 0.00001	< 0.00001	< 0.05
Fusiform	< 0.00001	< 0.0001	< 0.001	< 0.00001	< 0.001	< 0.01
Supramarginal	< 0.00001	< 0.0001	ns	< 0.00001	< 0.001	ns
Lateralorbitofrontal	< 0.00001	< 0.001	ns	< 0.00001	< 0.01	ns
Parsorbitalis	< 0.00001	< 0.001	ns	< 0.00001	< 0.001	ns
Bankssts	< 0.00001	< 0.0001	ns	< 0.00001	< 0.001	ns
Superiorfrontal	< 0.00001	< 0.001	< 0.05	< 0.00001	< 0.001	ns
Parsopercularis	< 0.00001	< 0.001	ns	< 0.00001	< 0.01	ns
Insula	< 0.00001	< 0.001	< 0.01	< 0.00001	< 0.001	< 0.01
Rostralanteriorcingulate	< 0.00001	< 0.01	< 0.05	< 0.00001	< 0.001	< 0.001
Isthmuscingulate	< 0.00001	< 0.01	ns	< 0.00001	< 0.001	ns
Inferiorparietal	< 0.00001	< 0.001	< 0.05	< 0.00001	< 0.001	ns
Transversetemporal	< 0.00001	< 0.001	ns	< 0.001	< 0.05	ns
Caudalanteriorcingulate	< 0.00001	< 0.01	ns	< 0.00001	< 0.01	ns
Parstriangularis	< 0.00001	< 0.01	< 0.05	< 0.00001	< 0.01	< 0.01
Rostralmiddlefrontal	< 0.00001	< 0.05	< 0.0001	< 0.00001	< 0.05	< 0.01
Caudalmiddlefrontal	< 0.00001	< 0.01	ns	< 0.00001	< 0.01	ns
Posteriorcingulate	< 0.00001	< 0.01	ns	< 0.00001	< 0.01	ns
Precuneus	< 0.00001	< 0.01	ns	< 0.00001	< 0.01	ns
Medialorbitofrontal	< 0.00001	< 0.05	ns	< 0.001	ns	ns
Precentral	< 0.00001	< 0.05	< 0.05	< 0.0001	< 0.05	ns
Frontalpole	< 0.0001	< 0.05	ns	< 0.01	ns	ns
Postcentral	< 0.01	ns	< 0.00001	< 0.01	ns	< 0.00001
Superiorparietal	< 0.01	ns	< 0.00001	< 0.01	ns	< 0.0001
Lateraloccipital	< 0.01	ns	< 0.00001	< 0.05	ns	< 0.00001
Lingual	< 0.05	ns	< 0.00001	< 0.01	ns	< 0.00001
Paracentral	< 0.05	ns	< 0.01	< 0.01	ns	< 0.01
Pericalcarine	ns	ns	< 0.00001	ns	ns	< 0.00001
Cuneus	ns	ns	< 0.00001	ns	ns	< 0.00001

*Two-way Student t test is used for univariate analysis with a significant level of 0.05 for p-value.

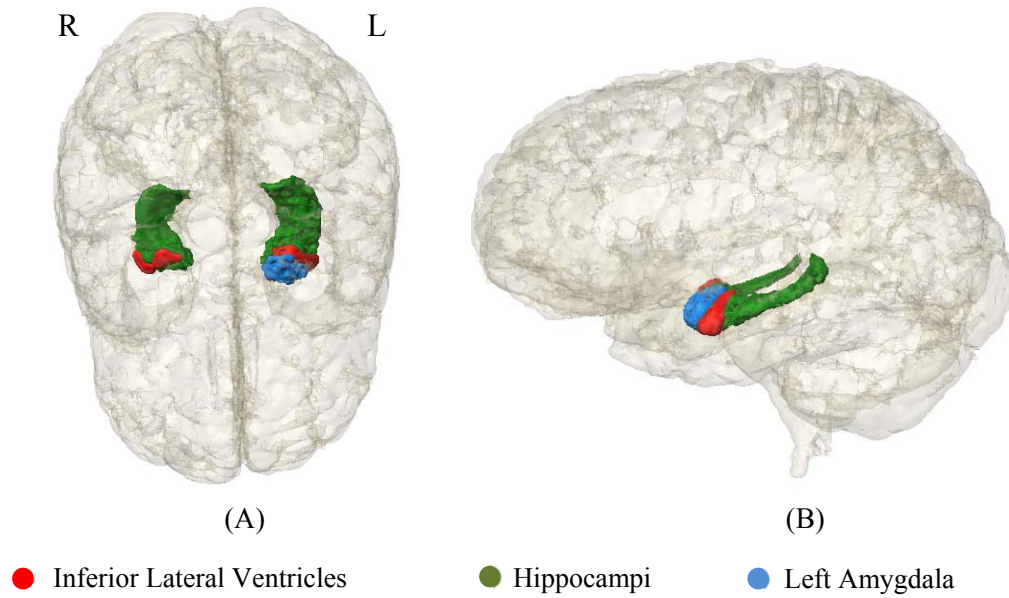


Figure 6: Representation of the top 5 significant subcortical volumes based on raw data in Table 12. (A) Superior view (B) Lateral view

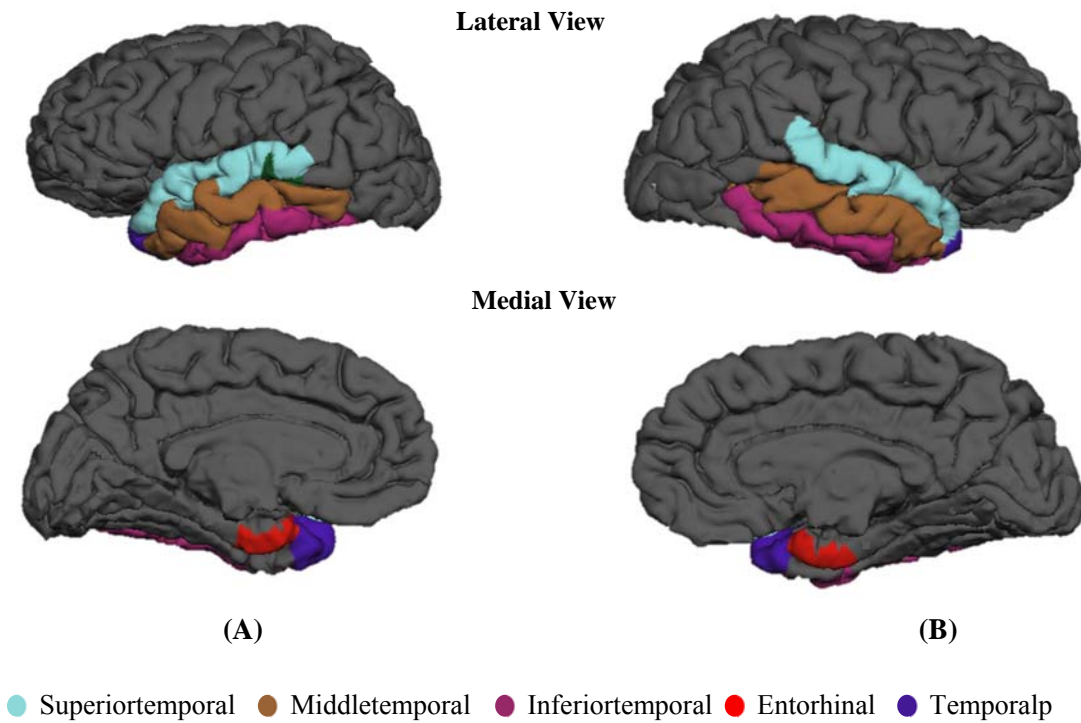


Figure 7: Representation of the top 5 significant cortical thickness based on raw data in Table 14. (A) Left hemisphere (B) Right hemisphere

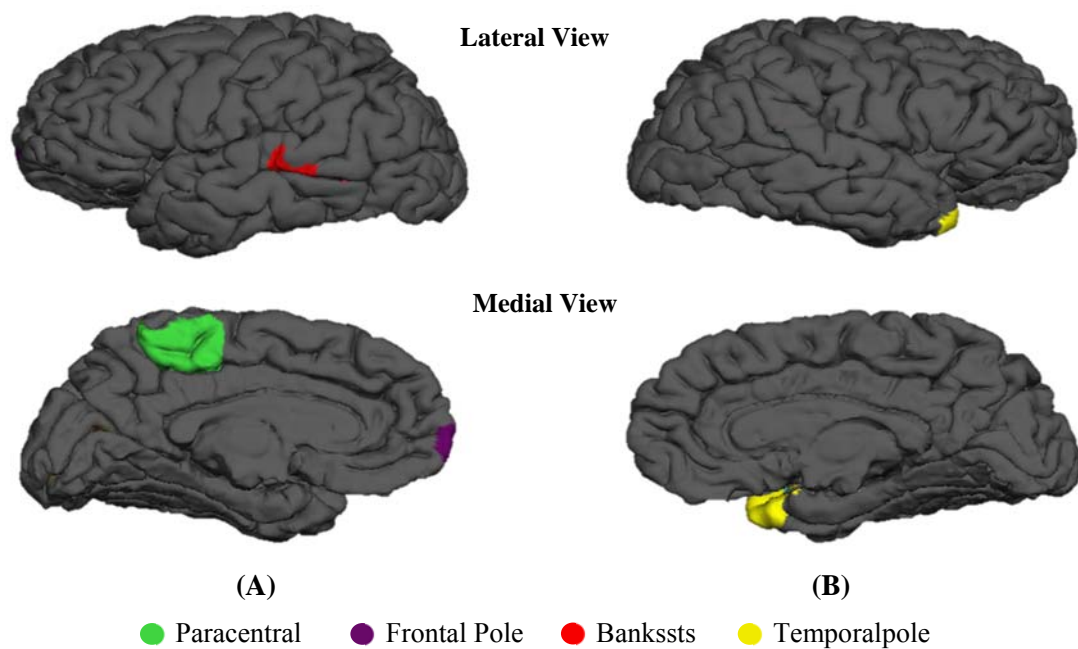


Figure 8: Representation of all significant surface area based on raw data in Table 13. (A) Left hemisphere (B) Right hemisphere

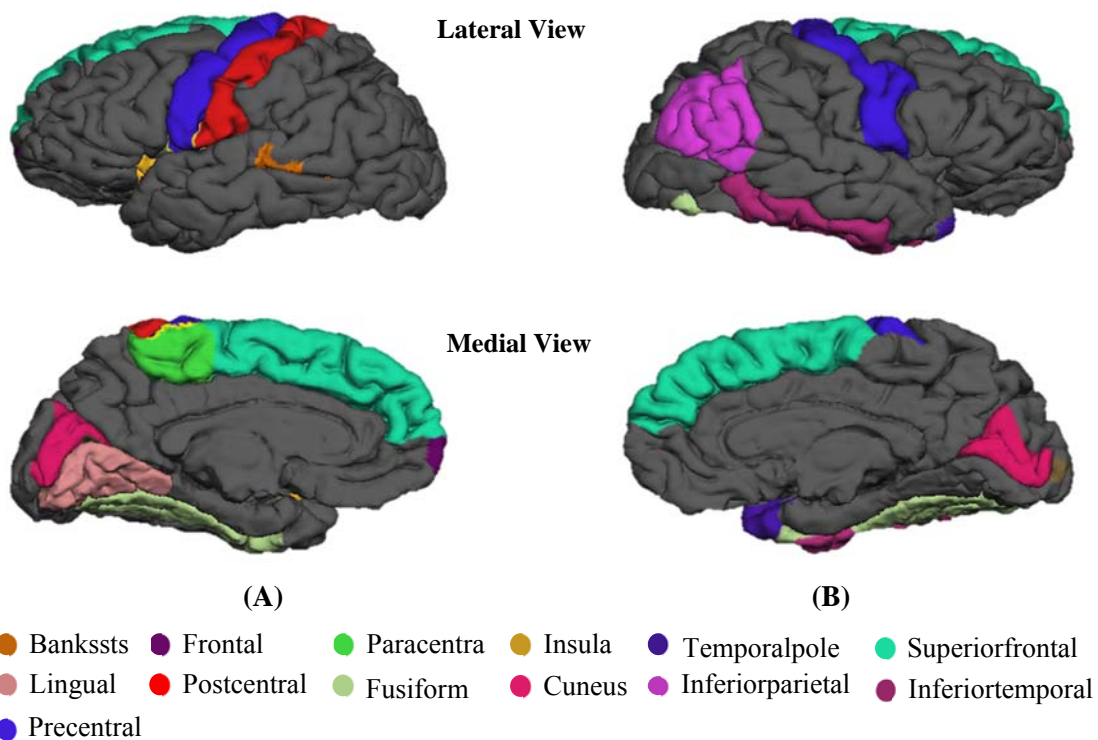


Figure 9: Representation of all significant surface area based on total-area normalized data on Table 13. (A) Left hemisphere (B) Right hemisphere

3.2.3 Spatial Distribution of Subjects Under the Best Model

Model of “MMSE + SV” without normalization gives the highest classification accuracy which utilizes the top 3 variables found within the model (i.e. MMSE, right-hippocampus volume and left-inferior-lateral-ventricle volume). One typical distribution of the data points for this classification model is plotted in Figure 10 to show the clustering characteristics of the data when MMSE and subcortical volumes are employed. Using this optimal decisional space, it can be observed that all the normal subjects are grouped into a very compact cluster, whereas AD subjects are more sparsely distributed in context of these dimensional parameters. This indicates the complex pattern of atrophy undergoing among the AD patients, which renders the classification task extremely difficult.

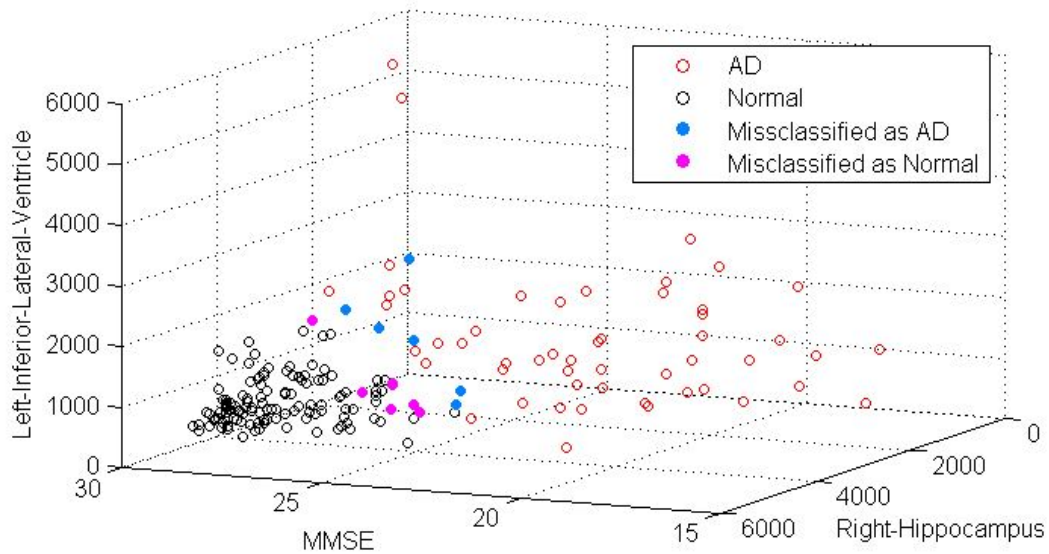


Figure 10: Representation of the whole dataset for the model of MMSE + subcortical volume, for a typical classification run under this model

3.2.4 Model Efficiency Estimation and Normalization

Variation in measures can come from many sources, including variation due to AD atrophy (σ_{ADa}^2), which is of primary interest for classification purposes, as well as other variation noise (σ_n^2) like individual difference in brain size, structure of brain regions, MRI measure error, region segmentation error, atrophy due to normal aging and resistance to brain atrophy (e.g. cognitive reserve). Generally, the total variance can be described as follows:

$$\sigma_{total}^2 = \sigma_{ADa}^2 + \sigma_n^2 \quad (3.1)$$

where σ_{total}^2 is the total variance of dataset, σ_{ADa}^2 stands for variance due to AD atrophy and σ_n^2 is the variance due to what is termed here as an overall source of noise. Also, discriminative power of a model depends on the amount of variance due to AD atrophy captured by the model used in contrast to the variance due to noise. A relevant term called discriminative power (Dp) can be estimated using Equation (3.2).

$$Dp = \frac{\widetilde{\sigma_{ADa}^2}}{\widetilde{\sigma_n^2}} \quad (3.2)$$

where $\widetilde{\sigma_{ADa}^2}$ is an estimate of the variance due to AD atrophy captured by the model, and $\widetilde{\sigma_n^2}$ stands for the estimated variance due to noise captured by the model.

Our results thus show that normalization in general does not enhance the classification performance significantly, which could be explained through Eq. 2 which shows that normalization does bring down correlated noise ($\widetilde{\sigma_n^2}$) experienced through brain size difference, but it also lowers down the correlated variance due to AD atrophy ($\widetilde{\sigma_{ADa}^2}$). A supporting finding of this assumption is that proportional volumes of the superior temporal cortex, expressed as a proportion of total cerebral volume was significantly different between females and males (Harasty, et al., 1997), which exemplifies the fact that normalization may be intrinsically biased. A similar finding by Barnes et al. is that normalization of all volumes by head size is not adequate due to their non-proportional relationship (Barnes, et al., 2010). Also Ross et al. found that males generally have a larger overall brain size than female, and males have larger cerebral cortical volumes than females except for left parietal (Carne, et al., 2006), thus normalization will at least bring in noise to the regions in left parietal as the regions in that area for males have a smaller size but normalized to a larger head size. However, the Dp value could still serve as a measure of a model's performance if relevant sources of the variance are known and are quantifiable, which is not the case in most practical scenarios.

CHAPTER 4

CLASSIFICATION AND PREDICTION BY PROPOSED EFFECTIVE MEAN INDICATOR

4.1 Background

The features extracted from biomarkers are often analyzed most often using various multivariate data analysis method such as principal component analysis (PCA) (López, et al., 2009; Nobili, et al., 2008), partial least square (PLS) (Higdon, et al., 2004; Ramírez, et al., 2010), and orthogonal partial least squares (OPLS) as exemplified in (Westman, et al., 2012a; Westman, et al., 2011b). The common objective of these techniques is to project the data into a decisional space where the total variance or variance related to class separation is maximized. Then linear or nonlinear using specific classifiers are then determined to delineate the populations or groups under study (Chaves, et al., 2009a; López, et al., 2009; Magnin, et al., 2009b; Zhou, et al., 2014d).

4.2 Methodology

4.2.1 Inclusion and Diagnostic Criteria

Only subjects from ADNI database with valid baseline MRI measures passing visual quality control (QC) evaluation were included in this study (as of February 2014 on ADNI website (<https://ida.loni.usc.edu/login.jsp>)). This has yielded a total of 543 qualified participants (CN = 139, EMCI = 220, LMCI = 108 and AD = 76). The

demographics and clinical characteristics of all the participants are as shown in Table 15.

Table 15: Demographics and characteristics of all participants

	CN	EMCI	LMCI	AD	p-value*
Number	139	220	108	76	-
Gender (female/male)	74/65	101/119	50/58	32/44	†0.194, 0.306, 0.153
Age	73.6 ± 6.0	69.9 ± 7.0	71.0 ± 7.6	75.1 ± 7.3	< 0.001
Years of education	16.4 ± 2.5	16.0 ± 2.7	16.4 ± 2.7	16.0 ± 2.6	0.343
ADAS-11	5.8 ± 3.1	7.8 ± 3.5	11.4 ± 4.9	21.3 ± 7.3	< 0.001
ADAS-13	9.1 ± 4.5	12.4 ± 5.2	18.3 ± 7.1	31.9 ± 8.5	< 0.001
MMSE	29.1 ± 1.1	28.5 ± 1.5	27.6 ± 1.8	22.9 ± 2.0	< 0.001

Data are presented as mean ± standard deviation where applicable. ADAS-11 = 11-item Alzheimer's Disease Assessment Scale-cognitive subscale, ADAS-13 = 13-item Alzheimer's Disease Assessment Scale-cognitive subscale. *Unless otherwise noted, one-way ANOVA was performed and p-values smaller than 0.05 are bolded. † Fisher's exact test was performed for CN vs. EMCI, CN vs. LMCI and CN vs. AD, respectively, and the corresponding p values are listed in the same order.

Diagnostic criteria for CN, EMCI, LMCI and AD are as follows:

CN subjects: MMSE scores between 24 and 30 (inclusive), a clinical dementia rating (CDR) score of 0, non-depressed, non-MCI, and non-demented; EMCI subjects: MMSE scores between 24 and 30 (inclusive), a subjective memory concern reported by subject, informant, or clinician, objective memory loss measured by education adjusted scores on delayed recall of one paragraph from Wechsler Memory Scale Logical Memory II (WMSLM II) (≥ 16 years: 9-11; 8-15 years: 5-9; 0-7 years: 3-6), a CDR of 0.5, absence of significant levels of impairment in other cognitive domains, essentially preserved activities of daily living, and an absence of dementia; LMCI subjects: Same as EMCI

with a difference only in objective memory loss measured by education adjusted scores on delayed recall of one paragraph from WMSLM II (≥ 16 years: ≤ 8 ; 8-15 years: ≤ 4 ; 0-7 years: ≤ 2); Mild AD Subjects: MMSE scores between 20-26 (inclusive), a CDR of 0.5 or 1.0, and meets NINCDS/ADRDA criteria for probable AD.

4.2.2 MRI Acquisition and Analysis

Original MRI scans were acquired from a variety of 3T scanners with protocols individualized for each scanner, as defined in MRI protocols from ADNI website (<http://adni.loni.usc.edu/methods/documents/mri-protocols/>). Cortical reconstruction and volumetric segmentation of MRI scans were performed by applying FreeSurfer 5.1 (<http://surfer.nmr.mgh.harvard.edu/>) to T1 weighted MRI image (MPR or IRSPGR) in NiFTI format which has been pre-processed by Mayo Clinic (gradient warping, scaling, B1 correction and N3 inhomogeneity correction). The detailed technical procedures were described in prior publications (Dale, et al., 1999b; Fischl and Dale, 2000; Fischl, et al., 2002a; Fischl, et al., 2004a). In this study, FreeSurfer generated 271 applicable MRI features (including 68 cortical volumes, 68 cortical thickness, 70 surface areas, 49 cortical volumes and 16 hippocampal subfields), and they were all reviewed by Mayo Clinic for quality control.

4.2.3 Preprocessing of Data

In order to remove potential biases from further analysis, all MRI measures are performed by linear regression tests against the affecting factors, namely age (as

demonstrated in Table 15) and ICV. The measures with resulting p values less than 0.05 are adjusted based on Equation (4.1) as per Chiang et al.(Chiang, et al., 2011).

$$M_a = M_{ua} - G_{ICV} \cdot (V_{sICV} - V_{mICV}) - G_{age} \cdot (A_s - A_m) \quad (4.1)$$

Where M_a is the adjusted measure, M_{ua} is the unadjusted measure, V_{sICV} and A_s are the subject ICV and age (in years), respectively; V_{mICV} and A_m are the means of ICV and age for all the CN subjects, respectively. The gradients G_{ICV} and G_{age} are derived by a region-specific linear regression against subject ICV and age based on all the participants. This regression method also has the advantage that the order of regression does not affect the results.

4.2.4 EMI Calculation

Theoretically, after correction, values of measures can be compared directly within the same measure, meaning that the value difference of the same measure across different subjects represents the true difference in that measure not due to head size difference or age. In addition, the measure should be the same for the CN group and the actual variance is supposedly due to measurement errors.

In this study, the mean of each measure are estimated by averaging each measure using the training controls. And then all measures are normalized to these means, with the expectation that values of the measure for controls will approximate 1, and values less than 1 will be a measure of the severity of regional atrophy, typical of AD pathology.

Student's t-test with Bonferroni correction for multiple comparisons is performed to rank all measures based on the corresponding p values. The mean of the top-ranked measures are calculated and defined as EMI, and a threshold is defined so as to optimally separate groups. Inclusion of measures such as ventricular volume, which increases in the presence of brain atrophy, and is a good indicator of the severity of the disease (Luxenberg, et al., 1987; Nestor, et al., 2008a; Zhou, et al., 2014d), requires the following correction:

$$M_{\text{mod}} = 2 - M \quad (4.2)$$

where the value of M is the MRI measure before modification and M_{mod} is the same measure after modification. (E.g., a value of 1.3 for the ventricles is converted to 0.7).

4.2.5 Validation and Incremental Error Analysis

Classifications for AD vs. CN, LMCI vs. CN and EMCI vs. CN are performed using the proposed EMI-based method. The experiment is based on 10-fold cross validation, meaning the whole dataset is equally divided into 10 subsets, each time one subset is selected as testing data and the remaining 9 subsets as training data until all subsets have been chosen as testing set once and only once. The two parameters needed for EMI calculation are both generated during the training process, namely the mean of each measure using the CN subjects and the determined top-ranked measures. In order to blind the testing set of this information, these two parameters are based on only the training data. The training process also finds an EMI threshold that separates the two compared

groups with optimal performance. Then this obtained threshold is applied to the EMIs of the testing subjects for prediction, and the performance in terms of accuracy, sensitivity and specificity are recorded. The same experiments are repeated 50 times to get the estimated average performance.

In this process, measures are deemed optimal if together they capture a comprehensive pattern of atrophy that maximizes classification accuracy and yet they do not introduce noise patterns that could negatively impact the classification outcome. Therefore, there should be a specific number of top-ranked features selected that will yield the highest performance instead of using all of them. To this end, an incremental error analysis is designed to explore the number of top-ranked measures that should be included in the EMI calculation so that the desired performance is achieved. The incremental error analysis is employed by firstly using only the 1st top-ranked measure, and then each time an additional next top-ranked measure is introduced in the EMI calculation until the last measure in the rank is included. The general structure of the proposed EMI-based method is schematically illustrated in Figure 11.

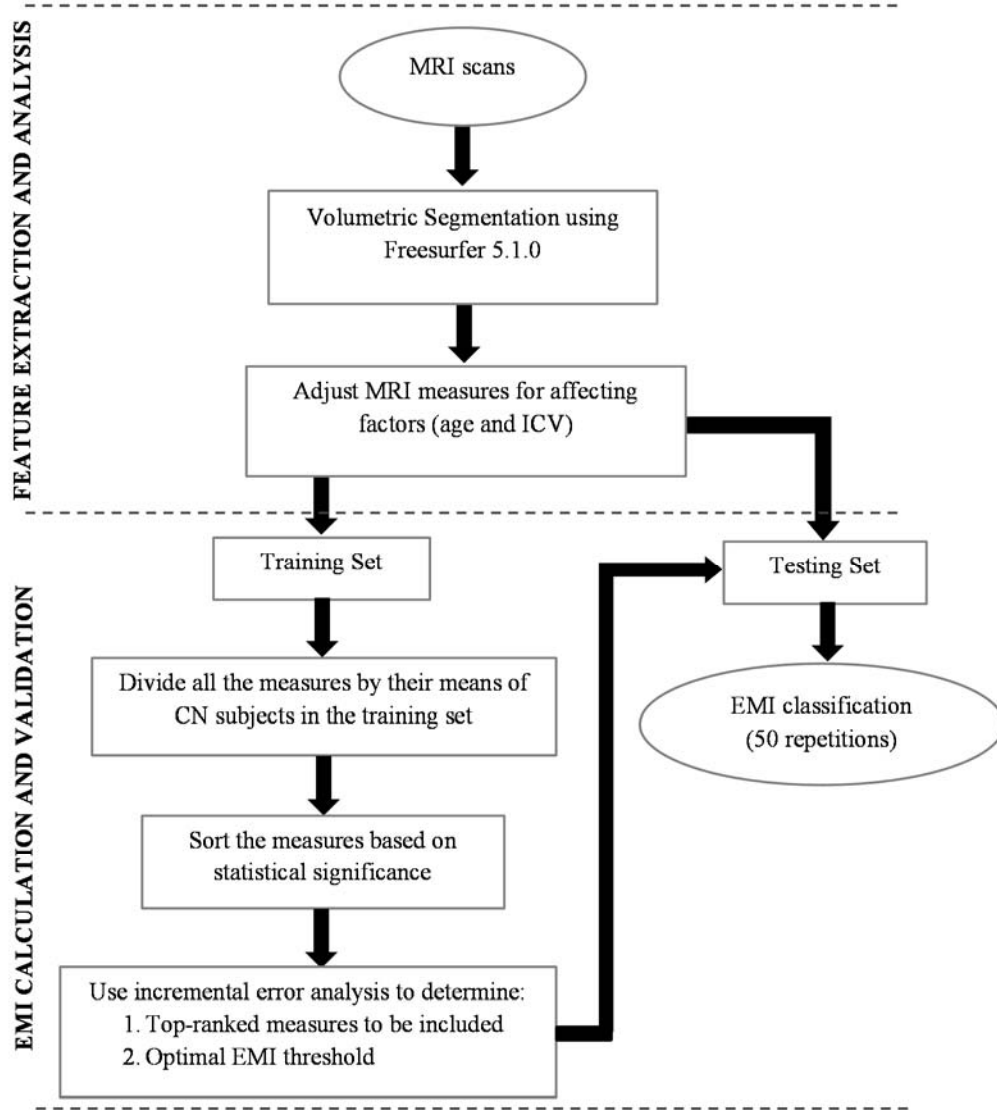


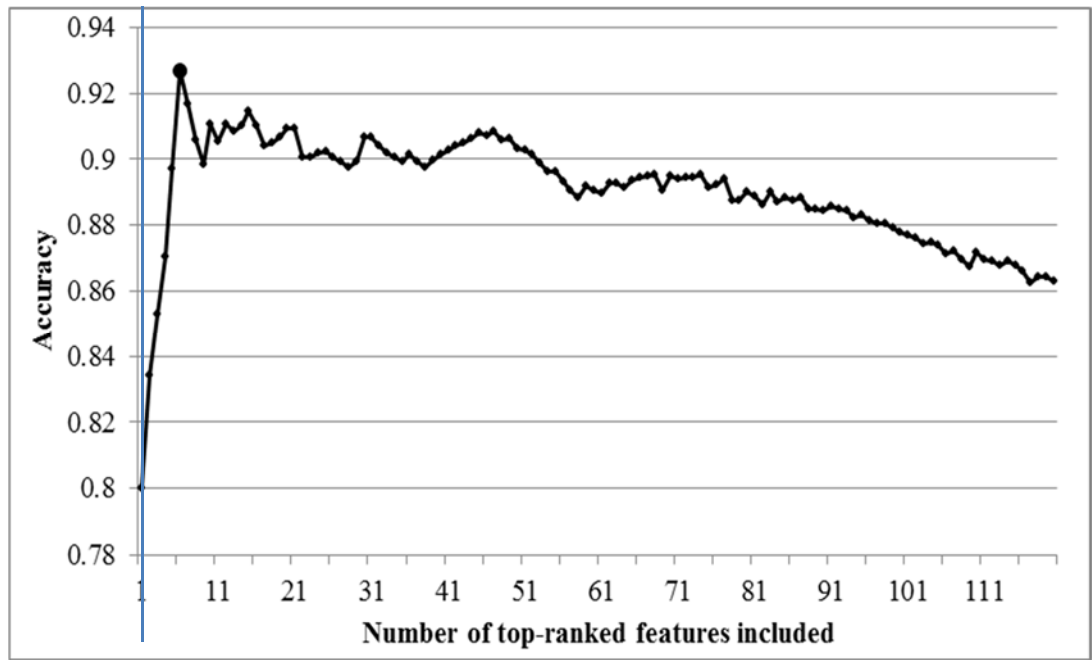
Figure 11: General structure of the proposed EMI-based method

4.3 Results

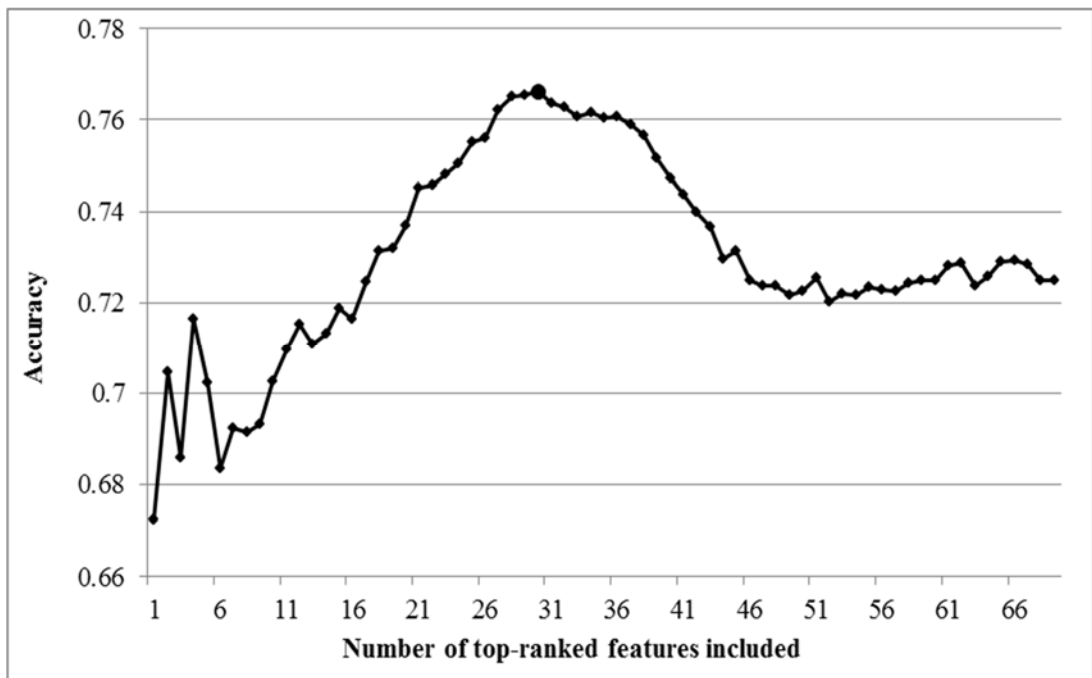
4.3.1 Incremental error analysis

In context to the structure shown in Figure 11, the incremental error analysis is designed to facilitate selection of the top-ranked measures that should be included in the

calculation of EMI and the corresponding EMI thresholds for prediction. The incremental error analysis is employed by starting from the top 1 measure and incrementally including the next-ranked measure for EMI calculation. The corresponding optimal separation thresholds for each inclusion of next-ranked measure are also recorded. Realizing the exhaustive nature of this process, and since many lower-ranked measures bear little information if any in the classification process, a Student's t-test with Bonferroni corrections for AD vs. CN, LMCI vs. CN and EMCI vs. CN using all subjects was used to select a total of 120 significant measures for AD, 69 for LMCI and 7 for EMCI. The accuracy as a function of the number of top-ranked measures included based on 50 independent 10-fold cross-validated experiments is plotted in Figure 12. As the measures are sorted in the training phase, which change during cross-validation, the rank of features may alter slightly from time to time. Therefore the incremental error analysis is designed to determine the number of top-ranked variables to be included. Figure 12 shows that for each classification type there is a unique highest point, which decides the number of top-ranked measures that should be included. Interestingly, for LMCI classification, EMI using 30 top-ranked measures yielded the highest performance in contrast with the top 6 for AD and the top 3 for EMCI. In addition, during the 50 repetitions of experiments for AD, the top 6 measures yielding the highest accuracy were observed 44 times, demonstrating extremely robust patterns of this AD model for EMI calculation. And 25 out of 50 times were observed for EMI using the 3 top-ranked measures.



(A)



(B)

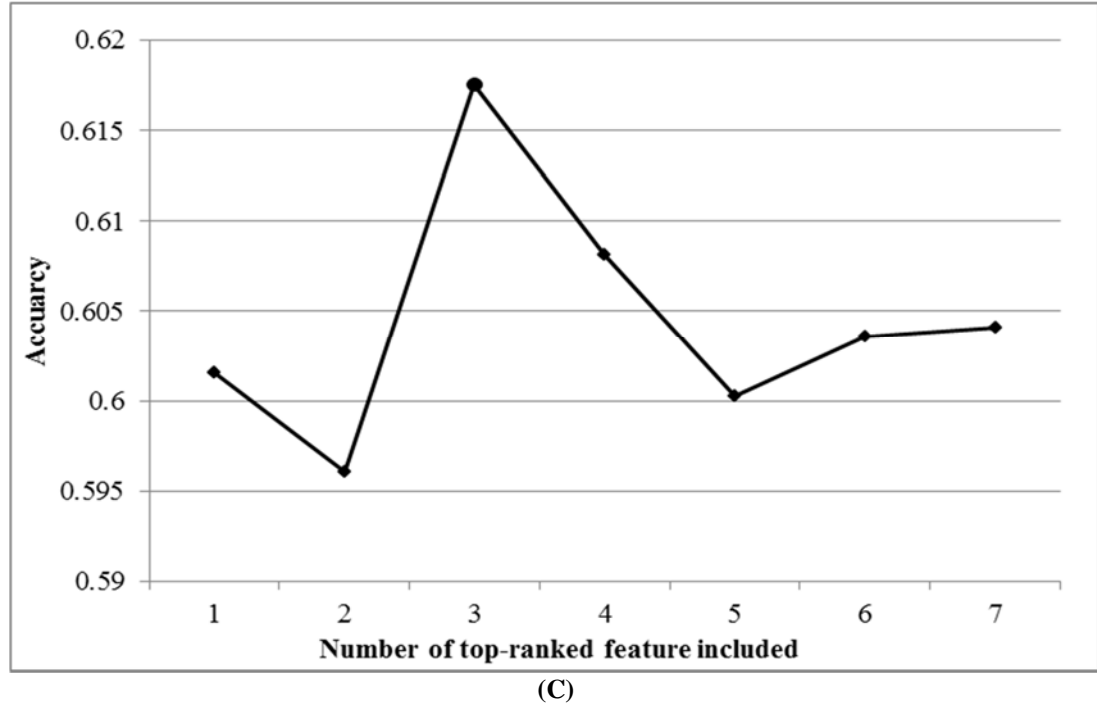


Figure 12: Incremental error analysis performance in terms of accuracy as a function of the number of top-ranked measures included for (A) AD vs. CN. (B) LMCI vs. CN. (C) EMCI vs. CN. The accuracy is the average obtained after 50 independent experiments based on 10-fold cross validation with the maximal point highlighted with a big dot.

The performances denoted by the maximal point in Figure 12 are also summarized in Table 16 in terms of accuracy, sensitivity and specificity. These results show that the EMI method is very efficient in differential diagnosis of AD with an accuracy of 92.7% for predicting AD. The numbers in parenthesis indicate the highest and lowest performance obtained during the 50 repetitions, which as can be observed are of small ranges of performance variation, confirming that the results are consistent.

Table 16: Classification performance of the EMI-based method with the determined top 6, 31 and 3 features, as denoted by the maximal points shown in Figure 12 for AD vs. CN, LMCI vs. CN and EMCI vs. CN, respectively

AD vs. CN			LMCI vs. CN			EMCI vs. CN		
ACC (%)	SEN (%)	SPE (%)	ACC (%)	SEN (%)	SPE (%)	ACC (%)	SEN (%)	SPE (%)
92.7	95.8	91.0	76.6	72.3	80.0	61.8	68.3	51.5
(91.2-93.1)	(93.6-96.3)	(89.2-91.4)	(74.8-78.5)	(69.4-74.3)	(77.6-83.5)	(58.5-65.5)	(62.3-72.7)	(41.8-60.3)

ACC = accuracy, SEN = sensitivity, SPE = specificity.

4.3.2 Histogram Plots, Boxplots and Scatter Plots of EMI

To investigate the separation efficiency of EMI for AD vs. CN, LMCI vs. CN and EMCI vs. CN, the EMI were calculated for all subjects based on the number of top-ranked measures found by the maximal points as shown in Figure 12. The descriptive statistics of the EMI are presented in Table 17. As expected, the mean of EMI for EMCI group is the highest and closest to that of CN, and AD is the lowest and furthest. Statistics tests also show that the mean difference of EMI between AD and CN is more significant than the other two cases as indicated by the p values in Table 16. It should also be noted that the standard deviation of CN group for different classification types varies according to the top-ranked measures that were used.

To better assess the separation efficiency of the EMI visually, histogram plots, boxplots and scatter plots of the EMI are shown in Figure 13 -Figure 15, respectively. Besides, these plots also have two lines depicting the largest and smallest thresholds resulting from the 50 independent training experiments. The two extreme thresholds form a range

of potential thresholds that may be obtained through the training process. The more compact is the range, the more consistent the results would be. In other words, these thresholds are like the separation boundaries of regular classifiers obtained repetitively using partial subjects. If the boundaries are similar or close, it means the classifier is less susceptible to population difference and is an indicator of robustness of the classifier.

Table 17: Descriptive statistics of the EMI

	AD vs. CN		LMCI vs. CN		EMCI vs. CN	
	CN	AD	CN	LMCI	CN	EMCI
Mean of EMI	1.00	0.76	1.00	0.86	1.00	0.94
S.D.	0.087	0.095	0.059	0.120	0.098	0.120
Kurtosis	3.24	2.63	2.98	2.75	2.74	4.33
P value by Student's t-test (Scale)	e^{-46}		e^{-22}		e^{-7}	

S.D. = standard deviation.

Figure 13 and Figure 15 jointly demonstrates that EMI-based method is efficient in discriminating different stages of the disease with very compact threshold range (for AD and LMCI in particular). In addition, the histogram in Figure 13 displays how the EMI of subjects are distributed in terms of sub-ranges. As expected, the EMI distribution of AD is more easily separable from CN distribution with less overlapping areas than LMCI and EMCI. Figure 14 graphically depicts the EMI values through their quartiles. It can be seen that the maximum of EMI among AD subjects is close to the median of EMI among CN subjects and the minimum of EMI in CN group excluding an outlier lies almost in the middle between the median and upper quartile of EMI among AD subjects. In

comparison, the EMI difference between LMCI (or EMCI) and CN groups are relatively smaller but still notable. Figure 15 shows the scatter plots of EMI against subjects' age as 2-D representations of the data. Even though class separation is solely based on the variation of EMI along the vertical axis, the age is used as the second dimension for visual appreciation. As can be observed, EMI separates AD from CN efficiently with any threshold in the determined range. Both the min and max thresholds separate LMCI from CN decently as well. As expected, prediction of EMCI is more challenging as shown in Figure 15C. It is worth noting that in Figure 15A, the EMI of AD subjects are quite stable around the threshold, even more so than the EMI of CN subjects near the threshold range. This observation is a strong indicator of the efficiency of EMI in identifying the pattern of AD and complies with the results in Table 16 that prediction of AD yielded a higher sensitivity of 95.8% than the specificity of 91.0%.

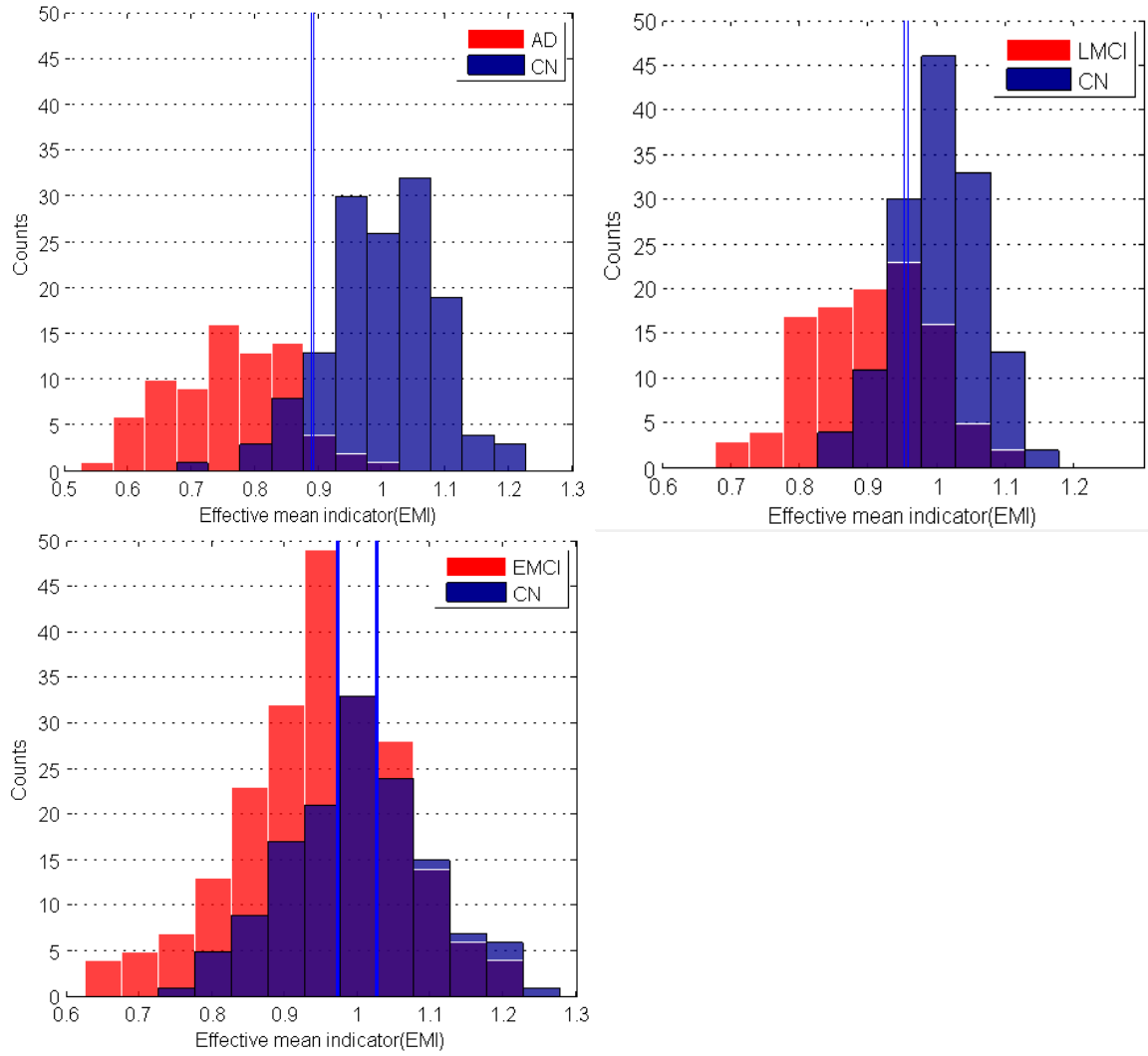


Figure 13: Histogram plots of the EMI using previously validated number of top-ranked measures for (top left) AD and CN. (top right) LMCI and CN. (bottom) EMCI and CN. In order to present the separation thresholds obtained from the training process of the 50 independent experiments, the maximum and minimum thresholds are plotted as blue lines.

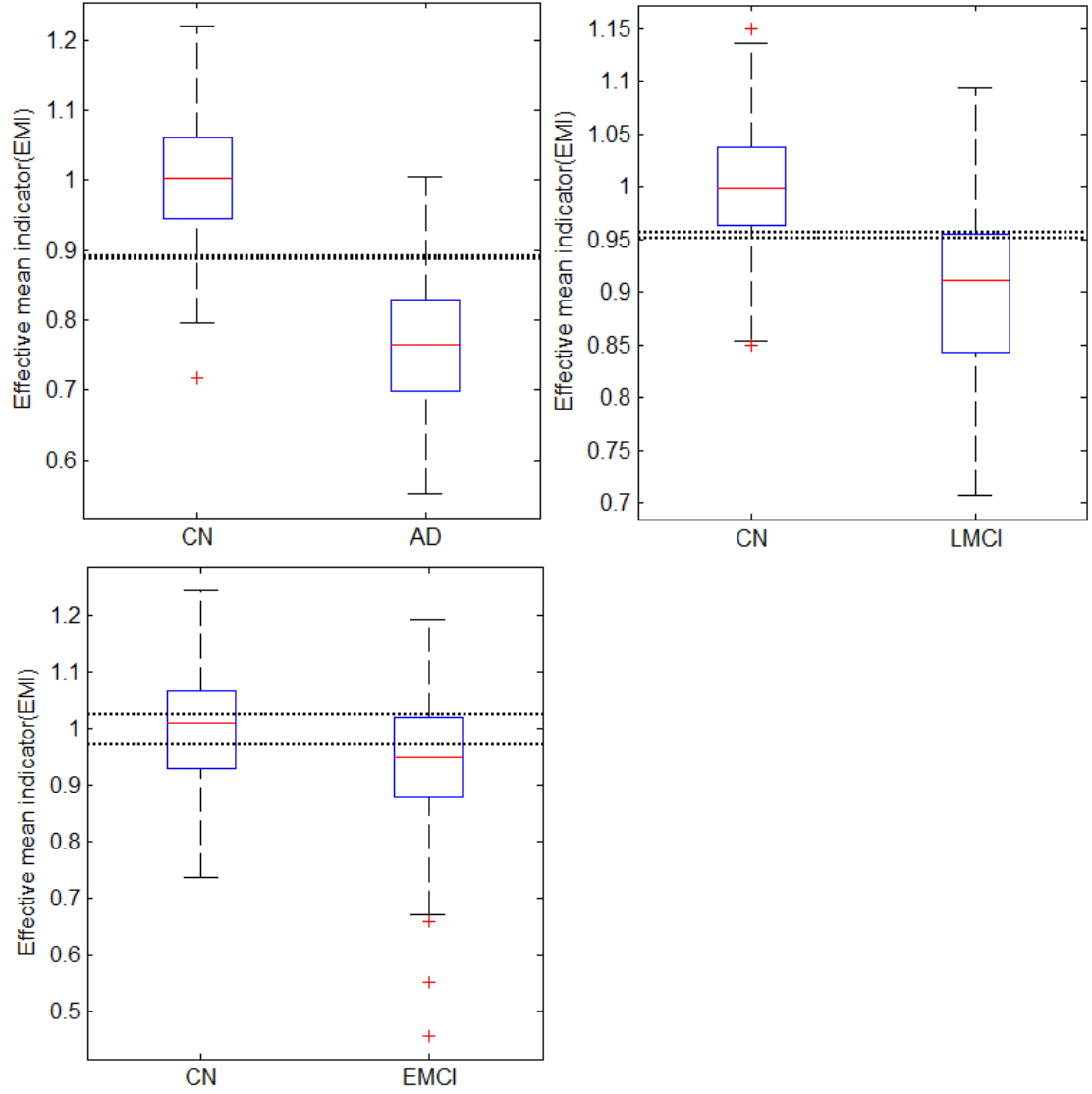
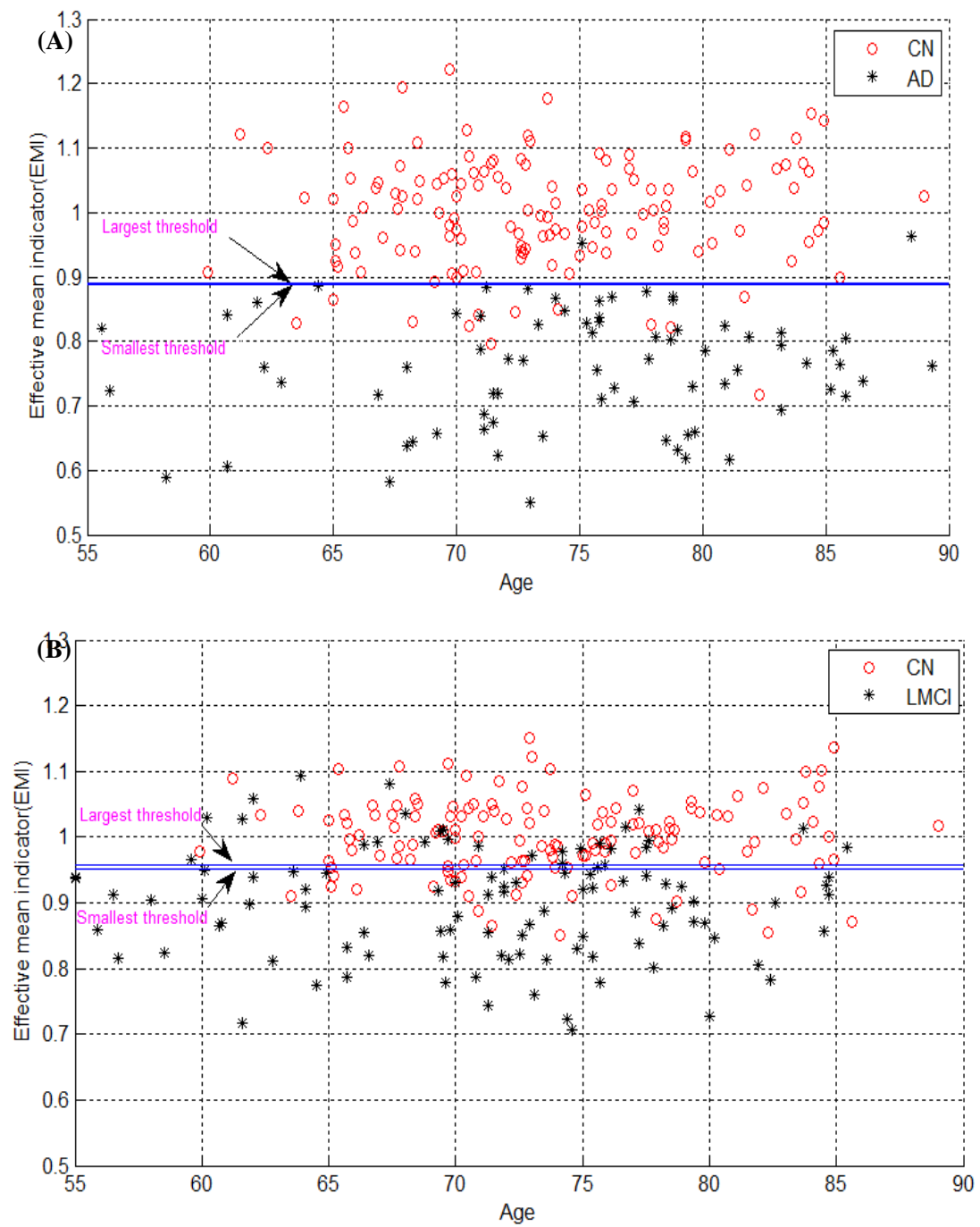


Figure 14: Boxplots of the EMI using previously validated number of top-ranked measures for (top left) AD and CN. (top right) LMCI and CN. (bottom) EMCI and CN. In order to present the separation thresholds obtained from the training process of the 50 independent experiments, the maximum and minimum thresholds are plotted as black dots.



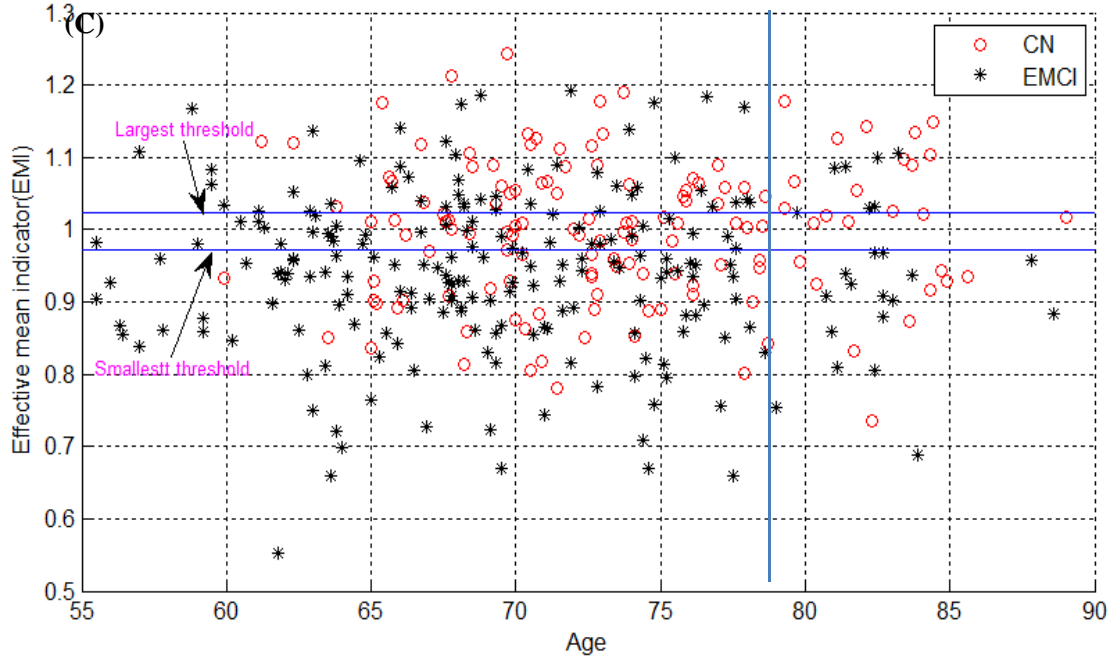


Figure 15: Scatter plots of the EMI using previously validated number of measures against age for (A) AD vs. CN. (B) LMCI vs. CN. (C) EMCI vs. CN. In order to present the separation thresholds obtained from the training process of the 50 independent experiments, the maximum and minimum thresholds are plotted as blue lines.

Furthermore, Figure 15 also enables us to examine if there is a potential relationship between subjects' age and EMI, which in this case showed no evidence of such a relationship. In addition, linear regression tests were also performed between subjects' age and EMI, showing no significant relationship between them.

4.4 Discussion

In this study, we presented a new and robust EMI-based method to predict different stages of AD, utilizing regional MRI measures normalized to the mean values based on CN subjects. The cumulative difference of these regional measures from the mean value of 1 is indicative of the severity of atrophy representative of Alzheimer pathology. We

have shown in this study that the proposed EMI-method demonstrates superior and consistent performance in predicting different stages of AD.

This new method has the advantage of simplicity and in converting MRI measures into values that can be directly added and subtracted across measures without the assumptions required of PCA and PLS methods (López, et al., 2009; Nobili, et al., 2008; Ramírez, et al., 2010; Westman, et al., 2011b). These assumptions in PCA and PLS make the data interpretation less intuitive, because the position and distance to the boundary of a testing subject in the decisional space may be difficult to quantify, especially if the decisional space is multi-dimensional. The advantage of EMI is that it provides a value that can be intuitively interpreted as a measure of the severity of atrophy typically found in AD, without requiring pretreatment of data using unit variance scaling (UVS) or other similar methods to balance the variance of measures. In methods, such as PCA and PLS, measures with large variance dominate and information-bearing measures with smaller variance may be overwhelmed. Also, pretreatment of data like UVS may also hinder the interpretation of the data while altering the variance of measures (van den Berg, et al., 2006) and therefore the information related to class separation may be compromised. The EMI-based method does not require such pretreatment of data so that all the raw information is maximally retained.

Along with its simplicity in implementation, the EMI-based method has also demonstrated a high level of predictability achieved using only the MRI biomarker as shown in Table 16. Recent studies have combined different biomarkers to explore their

complementary information (Davatzikos, et al., 2011; Fan, et al., 2008b; Hinrichs, et al., 2011; Westman, et al., 2012a; Zhang, et al., 2011) and found that the combined biomarkers has higher predictive power than using single biomarkers alone. Some of these studies are selected for a performance comparison with the proposed EMI-based method as shown in Table 18. The subjects in this study, derived from ADNI GO and ADNI 2, has two well-defined MCI stages (EMCI and LMCI) while ADNI 1 has only one as MCI stage.

The proposed EMI-based method is shown to be more effective in the prediction of AD than any other methods, even though the other methods used multiple biomarkers. The EMI method yielded a higher sensitivity than any other methods, thus improving the classification of AD. Despite the difference in grouping of MCI subjects, the performance of EMI method on LMCI classification is compared with the performance of MCI classification in a recent and frequently cited study by Zhang et al.(Zhang, et al., 2011).The proposed EMI-based method predicted LMCI using only MRI with an accuracy of 76.6% (sensitivity: 72.3% and specificity: 80.0%), which is comparable to the prediction of MCI using MRI, PET and CSF by Zhang et al.(Zhang, et al., 2011) with an accuracy of 76.4% (sensitivity: 81.8% and specificity: 66.0%).

In addition, the highest and lowest performances of the 50 independent experiments, as indicated in parenthesis in Table 18, show that EMI-based method has a consistent performance with an even smaller range of variation than the method by Zhang et al. Furthermore, as Westman et al. brought up the concern of cost in combining biomarkers

in a prior study (Westman, et al., 2012a), the proposed method is an efficient and cost-effective alternative for achieving high prediction performance without the extra cost for using other biomarkers other than MRI.

4.4.1. Flexibility in Balancing Between Performances

As indicated in Figure 11, there are 3 parameters generated by the training process, which are deemed essential for the prediction of the testing data. Among them, the first one is the mean of each measure calculated using the CN subjects in the training set, to which the testing data is normalized. The other two are generated by the incremental error analysis, namely the number of top-ranked measures and the optimal EMI threshold found during the training process. If a higher sensitivity for prediction is desired, a threshold can be selected such that less AD subjects are misclassified despite the possible cost of accuracy and/or specificity. In this study, the threshold in the training process was determined such that performance (P) as defined below reaches maximum.

$$P = \text{Accuracy} * 0.8 + \text{Sensitivity} * 0.2 + \text{Specificity} * 0.2 \quad (4.3)$$

Equation (4.3) assigns different weights to the 3 measures of performance depending on the 3 factors, which are really flexible depending on the demands on performance. For example, the factors as used in this study assign more weight to accuracy than sensitivity and specificity. This is another advantage of the EMI method that it allows for balancing between accuracy, sensitivity and specificity in a simple and convenient way. For example, if a more conservative diagnosis is required, a lower threshold will increase the

sensitivity to some extent, and such balancing of performance can hardly be implemented using other classifiers. Considering SVM classifiers as an example, which are frequently used in studies of AD (Klöppel, et al., 2008; López, et al., 2009; Zhang, et al., 2011; Zhou, et al., 2014c), when the decisional space is multidimensional, the decision boundary has to move in every single dimension according to the distribution of training data and there are many potential boundaries that may achieve the same goal such as a higher sensitivity. More importantly, it is hard to justify which new boundary to use to achieve the desired effect in prediction. Neither of these problems applies to the proposed EMI-based method, as it makes its decision in only one dimension. In other words, it integrates all the separation information into one dimension while achieving even higher prediction performance in comparison to other methods.

Table 18: Performance comparison of the proposed EMI-based method with other methods for predicting stages of Alzheimer's disease

<i>Methods</i>	<i>Subjects (AD/CN)</i>	<i>Repetitions (cross validation)</i>	<i>Biomarkers</i>	<i>Accuracy (%)</i>	<i>Sensitivity (%)</i>	<i>Specificity (%)</i>
(Hinrichs, et al., 2011)*	ADNI(48/66)	30 (10-fold)	MRI +PET MRI+PET+CSF+APOE +Cognitive Scores	87.6 92.4	78.9 86.7	93.8 96.6
(Zhang, et al., 2011)	ADNI(51/52)	10 (10-fold)	MRI MRI + PET + CSF	86.2 (82.9 – 89.0) 93.2 (89.0-96.5)	86.0 (82.7-88.7) 93.0 (88.7-96.3)	86.3 (83.1-89.1) 93.3 (89.1-96.6)
(Westman, et al., 2012a)	ADNI(111/96)	- (7-fold)	MRI MRI + CSF	87.0 (81.7-90.9) 91.8 (87.2-94.8)	83.3 (74.6-89.5) 88.5 (80.6-93.5)	90.1 (83.1-94.3) 94.6 (88.7-97.5)
Proposed method	ADNI (139/76)	50 (10-fold)	MRI	92.7 (91.2-93.1)	95.8 (93.6-96.3)	91.0 (89.2-91.4)

*The information of the highest and lowest performances is not available and thus left blank.

4.4.2. Future Potential of EMI Method

As mentioned before, only the CN subjects in the training process were used to compute the mean of each measure, which is then used to normalize all measures for both the training and testing data. And in this study, there are 125 CN subjects in the training set (10-fold cross validation). Even with the limited samples of CN subjects used in this study, the prediction performance is already high and stable. If more CN subjects are available, the resulting mean of each measure will be more accurate and it will also improve the experiment performance. In addition, another goal of the training process is to find the threshold that optimally separates the groups, which has also been shown to be highly convergent, especially for AD and LMCI prediction as displayed in Figure 13 Figure 15 with 2 almost overlapped straight lines) depicting minimum and maximum thresholds of the 50 independent experiments. The consistency of thresholds will again be expected to improve along with the refinement of the mean of each measure because they are found based on the subjects' EMI, which is closely related to the measurements.

Overall, more accurate mean of each measure in the future may yield more accurate EMI values and hence more consistent prediction results. Another merit of EMI-based method is that the value of EMI itself is very intuitive, and a good representative in the level of atrophy. This indicates its great potential in clinical trials by giving the physicians an estimated level of the disease with those quantifiable values. And such a value may also be potentially combined with some other screening test scores to generate an even more information-bearing composite for diagnosis.

CHAPTER 5

MULTIVARIATE ANALYSIS OF MRI AND PET (FDG AND 18F-AV-45)

5.1 Background

Subjects with AD suffer from progressive loss of cognitive functions, which are associated with regional brain atrophy (Fox, et al., 1996). As a result, magnetic resonance imaging (MRI) capable of capturing brain structure has been widely used to analyze structural change of the disease (Fox and Schott, 2004; Zhou, et al., 2014b). Regional atrophy in hippocampus and amygdala (Laakso, et al., 1996; Laakso, et al., 1995a; Thompson, et al., 2004a), cortical thinning (Dickerson, et al., 2009a) and ventricular enlargement (Thompson, et al., 2004a) were found significant related to anatomical change of AD. It's also been reported that extracellular beta-amyloid ($A\beta$) deposition (Hardy and Allsop, 1991a) and tau protein abnormalities (Mudher and Lovestone, 2002b) were the fundamental causes of the disease, which have made amyloid positron emission tomography (PET) imaging modality popular as it's able to evaluate amyloid plaque level and hypometabolism of the brain, which contain discriminative information of AD staging (Nordberg, et al., 2010b). There are three common PET imaging agents being used, Fluorine-18-fluorodeoxyglucose (FDG), Florbetapir 18F (AV-45) and Pittsburgh compound B (PiB). PiB and AV-45 work similarly as they both bind to $A\beta$. While AV-45 has much longer half-life time than PiB, which allows for more significant accumulation of the tracer in the brain of AD subjects (Wong, et al., 2010). Researches

using AV-45-PET have reported high $A\beta$ deposition among AD subjects in some specific areas, such as precuneus, frontal and temporal cortices (Choi, et al., 2009; Wong, et al., 2010). Recent studies using FDG-PET have reported significant correlation of dementia severity and FDG uptake level in defined areas, such as posterior cingulate, temporoparietal, prefrontal association cortex and temporal cortex (Herholz, et al., 2002; Piert, et al., 1996). FDG PET and AV-45 PET images are highly correlated while also providing complimentary information (Hsiao, et al., 2012).

Orthogonal partial least square to latent structures (OPLS) is a supervised multivariate data analysis method that has shown its efficiency in analyzing complex biological data (Bylesjo, et al., 2006; Ray, 2012; Trygg and Wold, 2002; Westman, et al., 2012c; Westman, et al., 2011b). For AD study, Westman et al. combined manual hippocampal volume measurements with automated regional and global volume measures to discriminate AD and MCI from controls. By comparing the discriminative powers of these features, they reported that OPLS showed great potential in the prediction AD and MCI and in the conversion from MCI to AD (Westman, et al., 2011b). They also compared and combined MRI data from the European AddNeuroMed and ADNI using OPLS method, and the results indicated that the two cohorts showed similar pattern of atrophy and predictive power (between 80 and 90%) (Westman, et al., 2011a). In addition, they combined CSF and MRI measures for classification of AD and MCI conversion using OPLS and showed that combinative power of them were better than MCI and CSF separately (Westman, et al., 2012c).

5.2 Methodology

5.2.1 Dataset

The data used in this article were downloaded from ADNI website (<https://ida.loni.usc.edu/login.jsp>). All recruited participants were between 55-90 (inclusive) years of age, had a reliable study partner, able to provide an independent evaluation of functioning, and speak either English or Spanish. They were willing and able to undergo all test procedures including neuroimaging and lumbar puncture and agreed to longitudinal follow up. Specific psychoactive medications were excluded.

5.2.2 Inclusion and diagnostic criteria

As this study aimed to find the discriminative power of features extracted from MRI and PET imaging modalities, only subjects with valid FDG and AV-45-PET scans and MRI measures passing visual quality control (QC) evaluation at baseline were included. This yielded a total of 524 subjects qualified for this study (CN = 137, EMCI = 214, LMCI = 103 and AD = 70) as of February 2014 on ADNI website (<https://ida.loni.usc.edu/login.jsp>). The demographics and clinical characteristics of all participants are as shown in Table 19.

Diagnostic criteria for CN, EMCI, LMCI and AD are based on ADNI protocol (online: http://www.adni-info.org/pdfs/adni_protocol_9_19_08.pdf). CN subjects: MMSE scores between 24 and 30 (inclusive), a CDR of 0, non-depressed, non-MCI, and non-demented; EMCI subjects: MMSE scores between 24 and 30 (inclusive), a subjective memory

concern reported by subject, informant, or clinician, objective memory loss measured by education adjusted scores on delayed recall of one paragraph from Wechsler Memory Scale Logical Memory II (WMSLM II) (≥ 16 years: 9-11; 8-15 years: 5-9; 0-7 years: 3-6), a CDR of 0.5, absence of significant levels of impairment in other cognitive domains, essentially preserved activities of daily living, and an absence of dementia; LMCI subjects: Same as EMCI with a difference only in objective memory loss measured by education adjusted scores on delayed recall of one paragraph from WMSLM II (≥ 16 years: ≤ 8 ; 8-15 years: ≤ 4 ; 0-7 years: ≤ 2); Mild AD Subjects: MMSE scores between 20 and 26 (inclusive), a CDR of 0.5 or 1.0, and meets NINCDS/ADRDA criteria for probable AD.

5.2.3 MRI and PET

Both MRI and PET scans were at baseline. MRI scans were acquired from a variety of 3T scanners with protocols individualized for each scanner, as defined in (<http://adni.loni.usc.edu/methods/documents/mri-protocols/>). Briefly, all PET images were acquired 30 to 60 minutes post-injection, co-registered, averaged, reoriented into a standard 160×160×96 voxel image grid, and smoothed to a uniform isotropic resolution of 8 mm full width of maximum. A detailed description of PET protocols and acquisition procedures can be found in: <http://adni.loni.usc.edu/methods/pet-analysis/pre-processing..>

Table 19: Demographics and characteristics of participants

	Training Set			Testing set				p-value*
	CN	EMCI	LMCI	AD	CN	EMCI	LMCI	
Number	70	70	70	70	67	144	33	-
Gender (female/male)	33/37	29/41	33/37	31/39	38/29	69/75	16/17	0.865†
Age	73.6 ± 6.6	70.6 ± 6.8	71.2 ± 7.4	75.0 ± 7.9	73.6 ± 5.4	69.6 ± 7.0	70.9 ± 8.5	< 0.001
Years of education	16.5 ± 2.4	15.5 ± 2.6	16.7 ± 2.7	15.9 ± 2.6	16.3 ± 2.6	16.3 ± 2.6	16.1 ± 2.8	0.192
MMSE	29.2 ± 1.0	28.5 ± 1.4	27.5 ± 1.8	22.8 ± 2.1	29.0 ± 1.3	28.4 ± 1.6	27.9 ± 1.9	< 0.001
ADAS-11	5.5 ± 2.6	7.9 ± 3.4	11.2 ± 4.9	21.2 ± 7.5	6.2 ± 3.5	7.8 ± 3.5	11.8 ± 5.1	< 0.001
ADAS-13	8.9 ± 4.0	12.1 ± 5.7	18.2 ± 7.2	31.8 ± 8.7	9.4 ± 5.0	12.5 ± 5.1	18.6 ± 7.3	< 0.001
FDG	6.6 ± 0.5	6.4 ± 0.6	6.2 ± 0.8	5.3 ± 0.7	6.6 ± 0.6	6.5 ± 0.6	6.5 ± 0.6	< 0.001
AV45	1.09 ± 0.19	1.17 ± 0.22	1.29 ± 0.23	1.41 ± 0.18	1.11 ± 0.17	1.16 ± 0.20	1.30 ± 0.24	< 0.001

Data are represented as mean ± standard deviation where applicable. ADAS-11 = 11-item Alzheimer’s Disease Assessment Scale-cognitive subscale, ADAS-13 = 13-item Alzheimer’s Disease Assessment Scale-cognitive subscale, FDG = PET with [18F]-fluorodeoxyglucose (FDG), AV45 = PET with 18F-AV-45 (florbetapir). *Unless otherwise noted, Student’s t-tests were performed for the listed factors between CN and AD in the training set and p-values were shown with those large than 0.05 deemed significant as bolded. †P value of Fisher’s exact test that was performed between CN and AD in the training set.

5.2.4 Image analysis

Cortical reconstruction and volumetric segmentation of MRI images were done by applying FreeSurfer 5.1 (available online: <http://surfer.nmr.mgh.harvard.edu/>) to T1 weighted MRI image (MPR or IRSPGR) in NiFTI format which has been pre-processed (gradient warping, scaling, B1 correction and N3 inhomogeneity correction) by Mayo Clinic. The detailed technical procedures were described in prior publications (Dale, et al., 1999a; Fischl and Dale, 2000; Fischl, et al., 2002b; Fischl, et al., 2004b). In this study, FreeSurfer generated 271 applicable MRI features which passed QC evaluation by Mayo Clinic as shown in Table 20, Table 21, Table 22, Table 23 and Table 24, with Table 24 showing all the 15 features extracted from PET scans including 8 features from AV-45-PET and 7 from FDG-PET. These tables also include the p values to indicate those features that are significant in separating the AD, EMCI and LMCI groups from CN subjects.

Data presented in Table 20 to Table 24 are p values of Student's t-test for left and right hemispheres separated by '|' where applicable. Regional or composite features shown significant between any comparing groups were bolded and with "ns" indicating as not significant,

Table 20: Statistical significance of cortical volumes by lobes

	Cortical volumes		
	EMCI	LMCI	AD
<i>Frontal Lobe</i>			
Superior Frontal	ns ns	<0.001 0.002	<0.001 <0.001
Rostral Middle Frontal	ns ns	ns ns	ns <0.001
Caudal Middle Frontal	ns ns	ns ns	0.005 0.002
Pars Opercularis	ns ns	ns ns	ns ns
Pars Triangularis	ns ns	ns ns	0.016 ns
Pars Orbitalis	ns ns	ns <0.001	0.002 <0.001
Lateral Orbitofrontal	ns ns	<0.001 0.005	<0.001 <0.001
Medial Orbitofrontal	ns ns	ns ns	ns <0.003
Precentral	ns ns	ns ns	ns ns
Paracentral	ns ns	ns ns	0.002 ns
Frontal Pole	ns ns	ns ns	ns ns
<i>Parietal Lobe</i>			
Superior Parietal	ns ns	ns ns	<0.001 <0.001
Inferior Parietal	ns ns	0.003 0.002	<0.001 <0.001
Supramarginal	ns ns	0.017 ns	<0.001 <0.001
Postcentral	ns ns	ns ns	0.019 ns
Precuneus	ns ns	ns ns	<0.001 <0.001
<i>Temporal Lobe</i>			
Superior Temporal	ns ns	<0.001 ns	<0.001 <0.001
Middle Temporal	ns ns	<0.001 <0.001	<0.001 <0.001
Inferior Temporal	0.005 0.034	<0.001 <0.001	<0.001 <0.001
Bankssts†	ns ns	0.014 0.012	<0.001 <0.001
Fusiform	ns ns	<0.001 <0.001	<0.001 <0.001
Transverse Temporal	ns ns	ns ns	ns ns
Entorhinal	ns ns	<0.001 <0.001	<0.001 <0.001
Temporal Pole	0.009 ns	0.002 <0.001	<0.001 <0.001
Parahippocampal	ns ns	0.004 0.014	<0.001 <0.001
<i>Occipital Lobe</i>			
Lateral Occipital	ns ns	ns ns	0.005 0.002
Lingual	ns ns	ns ns	0.020 ns
Cuneus	ns 0.010	ns ns	ns ns
Pericalcarine	ns ns	ns ns	ns ns
<i>Miscellaneous</i>			
Insula	0.017 ns	0.002 0.004	<0.001 <0.001
Caudal Anterior Cingulate	ns ns	ns ns	ns ns
Isthmus Cingulate	0.047 0.042	ns ns	0.005 ns
Posterior Cingulate	ns ns	ns 0.045	<0.001 <0.001
Rostral Anterior Cingulate	ns ns	ns <0.001	<0.001 ns
Hemisphere WM*	-	-	-
ICV*	ns ns	ns ns	ns ns

Table 21: Statistical significance of cortical thickness by lobes

	Cortical thickness		
	EMCI	LMCI	AD
<i>Frontal Lobe</i>			
Superior Frontal	ns ns	0.004 0.007	<0.001 <0.001
Rostral Middle Frontal	ns ns	ns ns	<0.001 ns
Caudal Middle Frontal	ns ns	0.012 ns	<0.001 <0.001
Pars Opercularis	ns ns	0.003 ns	0.002 0.017
Pars Triangularis	ns ns	ns ns	0.013 ns
Pars Orbitalis	ns ns	ns ns	ns <0.001
Lateral Orbitofrontal	ns ns	0.049 ns	<0.001 <0.001
Medial Orbitofrontal	0.024 ns	0.004 ns	0.015 0.012
Precentral	ns ns	ns ns	0.018 0.045
Paracentral	ns ns	ns ns	0.003 0.028
Frontal Pole	ns ns	ns ns	ns ns
<i>Parietal Lobe</i>			
Superior Parietal	ns ns	ns ns	<0.001 <0.001
Inferior Parietal	ns ns	<0.001 0.028	<0.001 <0.001
Supramarginal	ns ns	0.0059 ns	<0.001 0.002
Postcentral	ns ns	ns ns	0.023 0.009
Precuneus	ns ns	0.002 <0.001	<0.001 <0.001
<i>Temporal Lobe</i>			
Superior Temporal	ns ns	<0.001 0.002	<0.001 <0.001
Middle Temporal	ns ns	<0.001 0.002	<0.001 <0.001
Inferior Temporal	0.002 ns	<0.001 <0.001	<0.001 <0.001
Bankssts†	ns ns	0.015 ns	<0.001 <0.001
Fusiform	0.018 0.036	<0.001 0.002	<0.001 <0.001
Transverse Temporal	ns ns	ns ns	ns ns
Entorhinal	0.014 0.031	<0.001 <0.001	<0.001 <0.001
Temporal Pole	0.022 0.022	<0.001 <0.001	<0.001 <0.001
Parahippocampal	ns ns	ns ns	<0.001 <0.001
<i>Occipital Lobe</i>			
Lateral Occipital	ns ns	ns ns	0.006 0.005
Lingual	ns ns	ns ns	0.013 0.044
Cuneus	ns ns	ns ns	0.036 ns
Pericalcarine	ns ns	ns ns	ns ns
<i>Miscellaneous</i>			
Insula	ns ns	0.003 0.012	<0.001 <0.001
Caudal Anterior Cingulate	ns ns	ns ns	ns ns
Isthmus Cingulate	ns ns	0.011 0.009	<0.001 <0.001
Posterior Cingulate	ns ns	ns ns	<0.001 0.008
Rostral Anterior Cingulate	ns ns	ns ns	0.005 ns
Hemisphere WM*	-	-	-
ICV*	-	-	-

Table 22: Statistical significance of surface area by lobes

	Surface area		
	EMCI	LMCI	AD
Frontal Lobe			
Superior Frontal	ns ns	0.008 ns	0.015 ns
Rostral Middle Frontal	ns ns	ns 0.009	0.001 <0.001
Caudal Middle Frontal	ns ns	ns ns	ns ns
Pars Opercularis	ns ns	ns ns	ns ns
Pars Triangularis	ns ns	ns ns	ns ns
Pars Orbitalis	ns ns	0.037 0.027	0.006 0.021
Lateral Orbitofrontal	ns ns	0.008 ns	0.003 ns
Medial Orbitofrontal	0.015 ns	ns ns	ns ns
Precentral	ns ns	ns ns	ns ns
Paracentral	ns ns	ns ns	ns ns
Frontal Pole	ns ns	ns ns	ns ns
Parietal Lobe			
Superior Parietal	ns ns	ns ns	ns ns
Inferior Parietal	ns ns	ns 0.026	<0.001 <0.001
Supramarginal	ns ns	ns ns	0.016 0.034
Postcentral	ns ns	ns ns	ns ns
Precuneus	0.021 ns	ns ns	ns ns
Parietal Lobe			
Superior Temporal	ns ns	ns ns	0.022 ns
Middle Temporal	ns ns	0.011 <0.001	<0.001 <0.001
Inferior Temporal	ns ns	<0.001 0.007	<0.001 <0.001
Bankssts†	ns ns	0.047 0.040	0.002 <0.001
Fusiform	ns ns	0.008 ns	0.004 <0.001
Transverse Temporal	ns ns	ns ns	ns 0.044
Entorhinal	ns ns	ns 0.044	ns ns
Temporal Pole	ns ns	ns 0.013	ns ns
Parahippocampal	ns ns	0.018 0.023	<0.001 0.022
Occipital Lobe			
Lateral Occipital	ns ns	ns ns	ns ns
Lingual	ns ns	ns ns	0.039 ns
Cuneus	ns 0.021	ns ns	ns ns
Pericalcarine	ns ns	ns ns	ns ns
Miscellaneous			
Insula	ns ns	ns ns	0.018 ns
Caudal Anterior Cingulate	ns ns	ns ns	ns ns
Isthmus Cingulate	0.022 ns	ns 0.036	ns ns
Posterior Cingulate	ns ns	ns 0.048	ns ns
Rostral Anterior Cingulate	ns ns	ns ns	ns ns
Hemisphere WM*	ns ns	0.037 0.043	0.001 <0.001
ICV*	-	-	-

Table 23: Statistical significance of subcortical volumes by lobes

Subcortical volumes	EMCI	LMCI	AD
Brainstem	ns	ns	ns
CC Anterior	ns	ns	0.012
CC Central	ns	ns	0.010
CC Middle Anterior	ns	ns	0.034
CC Middle Posterior	ns	ns	ns
CC Posterior	ns	ns	0.022
Cortical GM	ns	<0.001	<0.001
Cortical WM	ns	0.030	0.008
CSF	ns	0.012	<0.001
Fourth Ventricle	ns	ns	ns
Non WM Hypo-Intensities	ns	ns	ns
Optic Chiasm	ns	ns	ns
Subcortical GM	ns	0.021	<0.001
Supra Tentorial	ns	0.002	<0.001
Third Ventricle	0.046	0.003	<0.001
Total GM	ns	<0.001	<0.001
WM Hypo-Intensities	0.005	0.008	<0.001
Accumbens Area	0.037 ns	0.004 <0.001	< 0.001 <0.001
Amygdala	0.002 0.007	<0.001 <0.001	< 0.001 < 0.001
Caudate	ns ns	ns ns	ns 0.049
Cerebellum Cortex	ns ns	ns ns	ns ns
Cerebellum WM	ns ns	ns ns	<0.001 ns
Choroid Plexus	ns ns	ns ns	0.018 0.008
Cortical GM	ns ns	<0.001 <0.001	<0.001 <0.001
Cortical WM	ns ns	0.027 0.035	0.006 0.013
Hippocampus	0.005 0.002	<0.001 <0.001	< 0.001 <0.001
Inferior Lateral Ventricle	0.002 0.021	<0.001 <0.001	<0.001 <0.001
Lateral Ventricle	ns ns	0.004 0.013	<0.001 <0.001
Pallidum	ns ns	0.027 ns	ns ns
Putamen	ns ns	ns ns	<0.001 <0.001
Thalamus	0.034 ns	0.038 ns	<0.001 <0.011
VentralDC	ns ns	ns ns	ns ns
Vessel	ns ns	ns ns	ns ns

Table 24: Statistical significance of hippocampus subfields and PET features

Hippocampus Subfields	EMCI	LMCI	AD
CA 1	ns ns	0.004 <0.001	<0.001 <0.001
CA 2-3	0.010 0.006	<0.001 <0.001	<0.001 <0.001
CA 4 (Dentate Gyrus)	0.023 0.006	<0.001 <0.001	<0.001 <0.001
Fimbria	ns ns	<0.001 0.016	<0.001 <0.001
Hippocampal Fissure	ns ns	ns ns	ns ns
Presubiculum	0.029 ns	<0.001 <0.001	<0.001 <0.001
Subiculum	ns ns	<0.001 <0.001	<0.001 <0.001
Tail	ns ns	<0.001 <0.001	<0.001 <0.001
AV-45 PET			
Cerebellum GM	ns	ns	ns
Whole cerebellum	ns	0.017	0.002
Brainstem	ns	0.002	<0.001
Frontal	ns	<0.001	<0.001
Cingulate	ns	0.002	<0.001
Parietal	ns	<0.001	<0.001
Temporal	ns	<0.001	<0.001
Summary of SUVR	0.017	<0.001	<0.001
FDG PET			
Cingulum Post (Bilateral)	ns	0.003	< 0.001
Temporal	0.011 ns	0.002 ns	<0.001 <0.001
Angular	0.016 ns	<0.001 0.009	<0.001 <0.001
FDG_sum	ns	<0.001	<0.001
Weighted average	ns	<0.001	<0.001

Processing of AV-45 PET scans was detailed in a prior article (Landau and Jagust, 2011a). Briefly, native-space MRI scans of subjects were used to define 4 cortical grey matter regions of interest (ROIs), i.e., frontal, anterior/posterior cingulate, lateral parietal and lateral temporal and 3 reference regions, i.e., cerebellar grey matter, whole

cerebellum and brainstem. Each florbetapir scan was co-registered to corresponding MRI and then the mean florbetapir uptake within the 7 cortical and reference regions was calculated as 7 regional AV-45 PET feature. Also a florbetapir composite feature (i.e. summary of standardized uptake value ratio (SUVR)) for each subject was created by non-weighted averaging across the 4 cortical regions and dividing this average by one of the reference regions (brainstem was used in this study).

Processing of FDG PET scans was detailed in another article(Landau and Jagust, 2011b) . Basically 5 MetaROIs (i.e. left and right Angular Gyrus, bilateral Posterior Cingular, left and right Inferior Temporal Gyrus) were identified by Landau et al. through well-defined procedures based on coordinates cited frequently in other FDG studies comparing AD, MCI, and CN(Landau, et al., 2011). In this article, 7 FDG features were considered, including FDG uptakes of the 5 MetaROIS and two other composite features, namely the sum and weighted average of FDG uptakes of the aforementioned 5 MetaROIs to represent FDG uptake level of the whole brain.

5.2.5 Preprocessing of data

All MRI measures were adjusted for age and ICV as per Equation (5.1) if the p value of the linear regression between the measure and the testing factor was smaller than 0.05. The aim was to remove potential biases from further analysis.

$$V_a = V_{ua} - G_{ICV} \cdot (V_{sICV} - V_{mICV}) - G_{AGE} \cdot (A_s - A_m) \quad (5.1)$$

where V_a is the adjusted measure, V_{ua} is the unadjusted measure, V_{sICV} and A_s are the subject ICV and age (years), respectively; V_{mICV} and A_m are the corresponding means for all the controls. The gradients G_{ICV} and G_{AGE} were derived by a region-specific linear regression against subject ICV and age of all the participants. As per Chiang et al. (Chiang, et al., 2011), the above regression also has the advantage that the regression order of age and ICV does not affect the regression result. Then all data was processed by mean centering and unit variance scaling.

5.2.6 Multivariate data analysis

The aforementioned 286 features were used as inputs to OPLS (Bylesjo, et al., 2006; Trygg and Wold, 2002), a supervised multivariate data analysis method comes with the software package SIMCA-P (version 11.5, Umetrics AB, Umea, Sweden). OPLS removes variation from descriptor variables that is not related to group separation and the information related to class separation is found in the predictive component (Trygg and Wold, 2002).

The predictive power of OPLS model for separating two groups is found in $Q_2(Y)$ and is defined as follows:

$$Q_2(Y) = 1 - \text{PRESS} / \text{SSY}_{\text{tot.corr}} \quad (5.2)$$

where PRESS (predictive residual sum of squares) = $\sum (y_{\text{actual}} - y_{\text{predicted}})^2$, is the squared differences between observed and predicted Y-values, and $\text{SSY}_{\text{tot.corr}}$ represents

the total variation of the Y variable (diagnosis) after scaling and mean centering (Eriksson, et al., 2006). $Q_2(Y)$ denotes the predictive power resulting from a 7-fold cross validation (by default). This procedure is repeated until every observation has been kept out once and only once. In this type of model, a $Q_2(Y)$ value larger than 0.5 is regarded as good (Eriksson, et al., 2006).

Variable influence in the projection (VIP) reflects significance of variables both with respect to Y (diagnosis) and X (the features included). It summarizes the overall contribution of each X-variable, summed over all other components and weighted according to the Y variation accounted by each component (Eriksson, et al., 2006; Galindo - Prieto, et al., 2014; Guyon and Elisseeff, 2003). And VIP has been shown to be critically important in selecting the significant variables (Ray, 2012; Weljie, et al., 2007). Variables with a VIP score larger than 1 are deemed significant as the average of squared VIP scores is equal to 1 (Eriksson, et al., 1995; Guyon and Elisseeff, 2003; Ray, 2012; Weljie, et al., 2007).

Multivariate analysis was done for the following two groups of models.

Full models: They included 7 single models and 3 hierarchical models for AD vs. CN, LMCI vs. CN and EMCI vs. CN. Single models used one of the 5 sets of MRI measures (i.e. subcortical volumes, cortical volumes, cortical thickness average, surface area and hippocampus subfields) or features from one of the PET (AV-45 or FDG) scans to explore discriminative power of different types of measures within the imaging

modalities. Three hierarchical models included one with all 5 sets of MRI measures, another one with two PET measures and a third one with all of them combined.

Feature-selected models: The aim of feature selection was to investigate if proper exclusion of less significant features would reduce inconsistent noisy patterns, and therefore enhance predictability. To this end, two feature selection methods were applied to differentiate EMCI, LMCI and AD from CN. The first scheme was commonly used feature selection technique that only those statistically significant features with a p value less than 0.05 were selected for each classification type. The other one was implemented iteratively based on VIP scores. At first, variables with VIP scores larger than 1 were selected from full models to create feature-selected models with one iteration of feature selection. Then feature-selected models with 2 iterations of feature selection included variables with VIP larger than 1 in the previous feature-selected models with one iteration. The same procedure was repeated to create more feature-selected models until the model reported a zero predictive power characterized by a $Q_2(Y)$ value of 0. The advantage of the VIP-based feature selection technique over the p value-based technique is that VIP score of the variable varies depending on the other variables included in the model, making it possible to iteratively exclude less significant variables from the model and explore predictive power of selected features.

5.2.7 Validation with external test set

Due to stringent inclusion criteria of subjects (i.e. subjects must have eligible MRI, valid FDG and AV-45 PET scans), among the four comparing groups, AD group has the

smallest sample size with 70 qualified subjects. In order to maintain high statistical power and avoid bias due to sample size discrepancy between groups during the training process, 70 subjects were randomly selected from the other three groups (CN, EMCI and LMCI) to match with the size of AD group, and the remaining subjects were defined as test set as shown earlier in Table 19.

Best full models and feature-selected models of LMCI vs. CN and EMCI vs. CN were validated using the combined LMCI-CN and EMCI-CN groups, respectively. To explore how best AD models recognize EMCI and LMCI subjects from CN, the best models of LMCI and EMCI were also cross validated with AD and CN subjects in the test set.

5.3 Results

5.3.1 Significance of features in separating groups

To interpret how the features may potentially contribute to separation, Student's t-test was performed for each feature between AD (EMCI or LMCI) and CN and the p values were as shown in Table 20 to Table 24. As expected, more variables showing significant difference were observed for discriminating AD from CN, and more MRI features of cortical volumes and cortical thickness were found significant than for surface area.

In terms of lobes, most significant regions were seen in temporal lobe with the least seen in occipital lobe. Almost all hippocampal subfields were shown significant for LMCI and AD except for hippocampal fissure. In particular, considering hemispheric symmetry,

most hemispheric measures showed similarity between hemispheres and asymmetric measures were mostly significant for the left hemisphere. Interestingly, SUVR across the whole brain, (defined as “summary of SUVR” in Table 24) was the only AV-45 PET measure that was shown to be significant for separating any stages of AD from CN. Besides, volumes of inferior temporal, amygdala, hippocampus, inferior lateral ventricle, and cortical thickness of fusiform and temporal were also shown significant between any of the compared groups.

5.3.2 Predictive power of OPLS models

Predictive power of full models: Predictive power ($Q_2(Y)$) for the aforementioned full models without feature selection is summarized in Table 25, which shows that combining MRI and PET features had the highest predictive power for models predicting LMCI and AD groups from the CN group, while using cortical thickness alone yielded a higher $Q_2(Y)$ for EMCI than using any measures, even when all features were combined.

Separation efficiency of the best full models during the training process could also be visualized with scatter plots as illustrated in Figure 16. Perfect separation of AD from CN using all features is shown in Figure 16A with a high $Q_2(Y)$ of 0.721. Moreover, Figure 16B and Figure 16C show that the models are efficient in separating LMCI and EMCI from CN as well.

Predictive power of feature-selected models: The two feature selection schemes using VIP scores and p values were applied to all features for EMCI, LMCI and AD against CN.

In addition, they were also applied to cortical thickness for EMCI prediction as cortical thickness was shown previously as the best full model in predicting EMCI. Predictive power of feature-selected models iteratively created with VIP function of SIMCA software are given in Figure 17 with black curves, of which the first point represents the predictive power of model with all features and the other points indicate predictive power of VIP-based feature-selected models with different numbers of iterations. Predictive power of p value-based feature-selected models was also shown in Figure 17 with straight or dotted lines. It could be seen that appropriate feature selection boosted predictive power as the linear segments kept rising up to a point, then began to drop as less and less features were selected in this iterative process. The results also showed that feature selection based on VIP scores offered more flexibility in selecting the best models with a peak predictive power higher than that obtained using statistical feature selection method.

In terms of numbers of iteration, Figure 17 indicates that for LMCI and AD prediction, 2 to 3 iterations of VIP-based feature selection yielded the maximal predictive power, while for EMCI prediction, no matter whether using all features or cortical thickness only, 1 iteration of VIP-based feature selection could already achieve the highest predictability.

Table 25: Summary of predictive power ($Q_2(Y)$) for all full models

Models		EMCI	LMCI	AD
MRI	Subcortical volume	N/A	0.188	0.585
	Cortical volume	N/A	0.257	0.528
	Cortical thickness	0.108	0.154	0.538
	Surface area	N/A	0.040	0.202
	Hippocampus Subfields	0.029	0.277	0.547
	<i>Combined</i>	<i>N/A</i>	<i>0.282</i>	<i>0.645</i>
PET	18F-AV-45	0.076	0.227	0.518
	FDG	0.038	0.055	0.512
	<i>Combined</i>	<i>0.093</i>	<i>0.229</i>	<i>0.636</i>
<i>MRI+PET</i>		<i>0.008</i>	<i>0.294</i>	<i>0.721</i>

N/A = Models had 0 predictive power ($Q_2(Y)$), FDG = PET with [18F]-fluorodeoxyglucose, AV45 = PET with 18F-AV-45 (florbetapir). The hierarchical models were italicized and highest $Q_2(Y)$ for three classification types were bolded.

5.3.3 Model validation with external test set

The best full models of EMCI and LMCI, shown as bolded in Table 25, were validated with the combined EMCI-CN and LMCI-CN groups, respectively. The same experiments were also done for best feature-selected models as indicated by the peaks of the curves in Figure 17. Besides, in order to investigate how models well trained with AD pattern would recognize and classify EMCI and LMCI subjects from controls in the test set, the best AD full models and best AD feature-selected models were also cross validated with EMCI and LMCI subjects against controls using external test set.

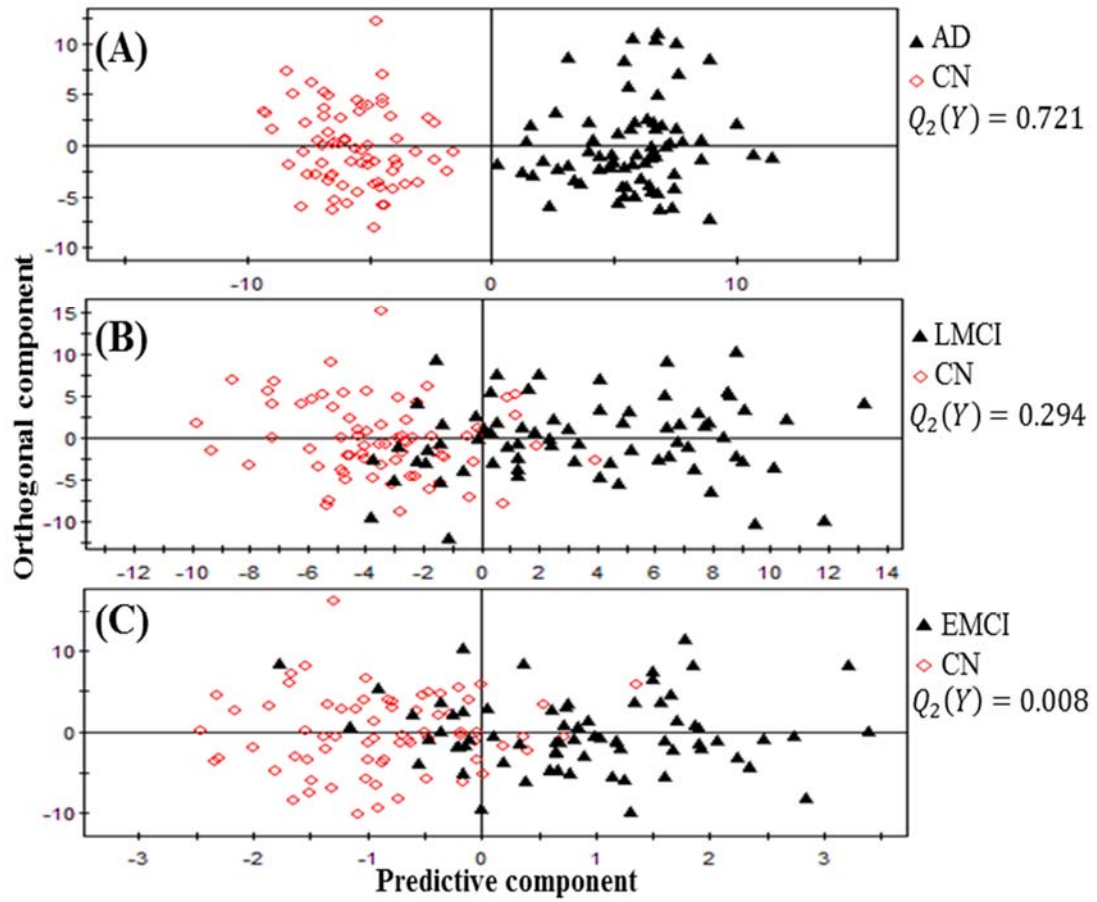


Figure 16: Scatter plots of best full models as indicated in Table 25 for separation between (A) AD and CN. (B) LMCI and CN (C) EMCI and AD.

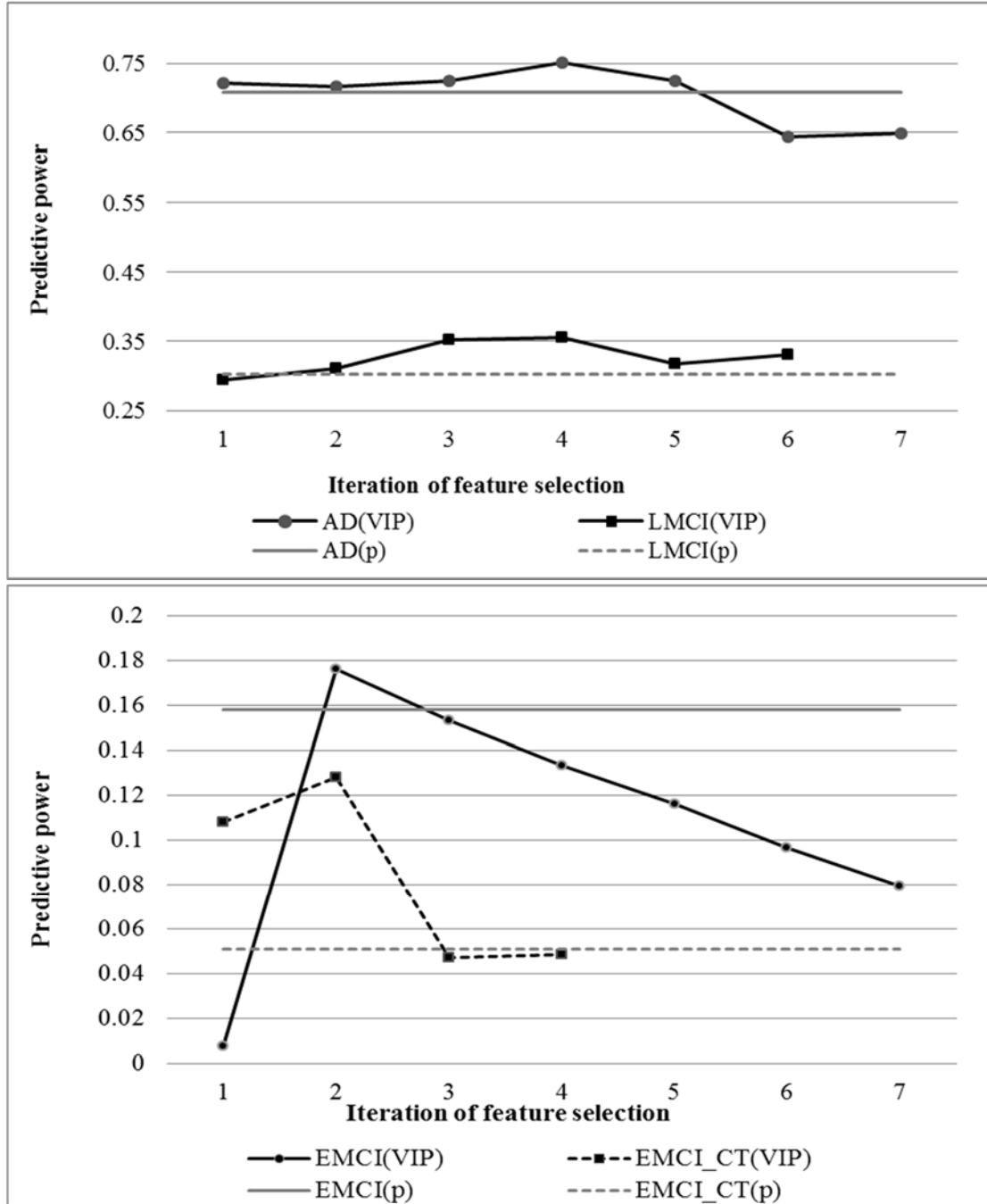


Figure 17: Predictive power ($Q_2(Y)$) of feature-selected models with p representing feature-selected models based on p values, and VIP denoting feature-selected models based on VIP scores. All of the feature selection techniques were applied to all features except for EMCI_CT, which was applied only to cortical thickness as it was shown previously to have the highest predictive power of EMCI: (top) predictive power of p value-based and VIP-based feature-selected models of AD and LMCI; (bottom) Predictive power of p value-based and VIP-based feature-selected models of EMCI using all features or cortical thickness only.

Scatter plots of validation for the best full models and best feature-selected models were shown in Figure 18 and Figure 19, respectively. Classification performance in terms of accuracy, sensitivity and specificity were summarized in Table 26. Comparing Figure 16C with Figure 18A - Figure 19A, it could be seen that neither best full model nor best feature-selected model of EMCI separated the testing EMCI from CN as well as it did during the training process, displaying poor generalization. However, this was not the case for best LMCI models. Figure 18C - Figure 19C and Figure 16B illustrate good generalization of LMCI models as LMCI subjects in the test set were classified nearly as accurately as it did during the training process. Figure 18B - Figure 19B and Figure 18D - Figure 19D show that AD models are very efficient in identifying controls, with Table 26 showing high specificity of more than 95%, which are, as expected, higher than the best EMCI or LMCI models, with LMCI specificity value being higher.

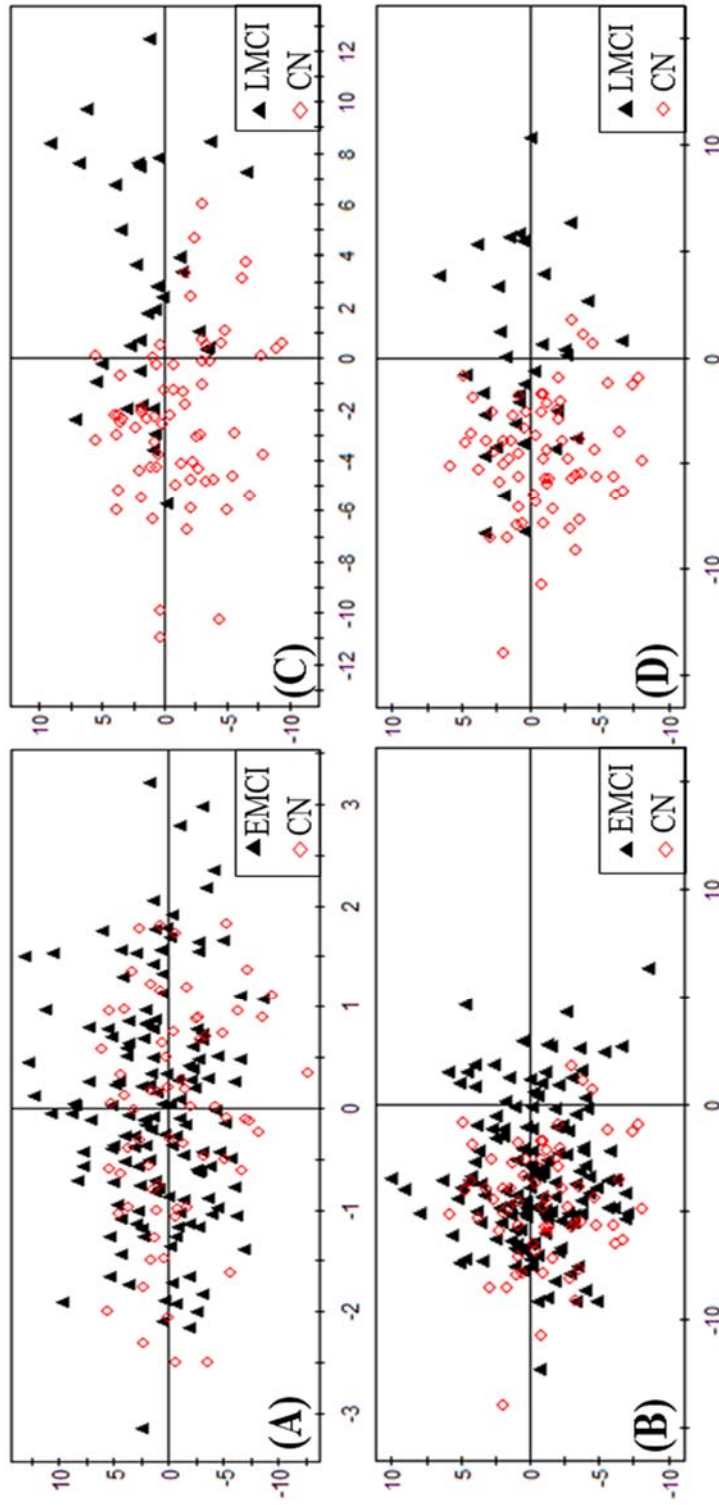


Figure 18: Scatter plots of best full models validated with external test set. (A) Testing best EMCI model with external EMCI and CN subjects. (B) Testing best AD model with external EMCI and CN subjects. (C) Testing best LMCI model with external LMCI and CN subjects. (D) Testing best AD model with external LMCI and CN subjects.

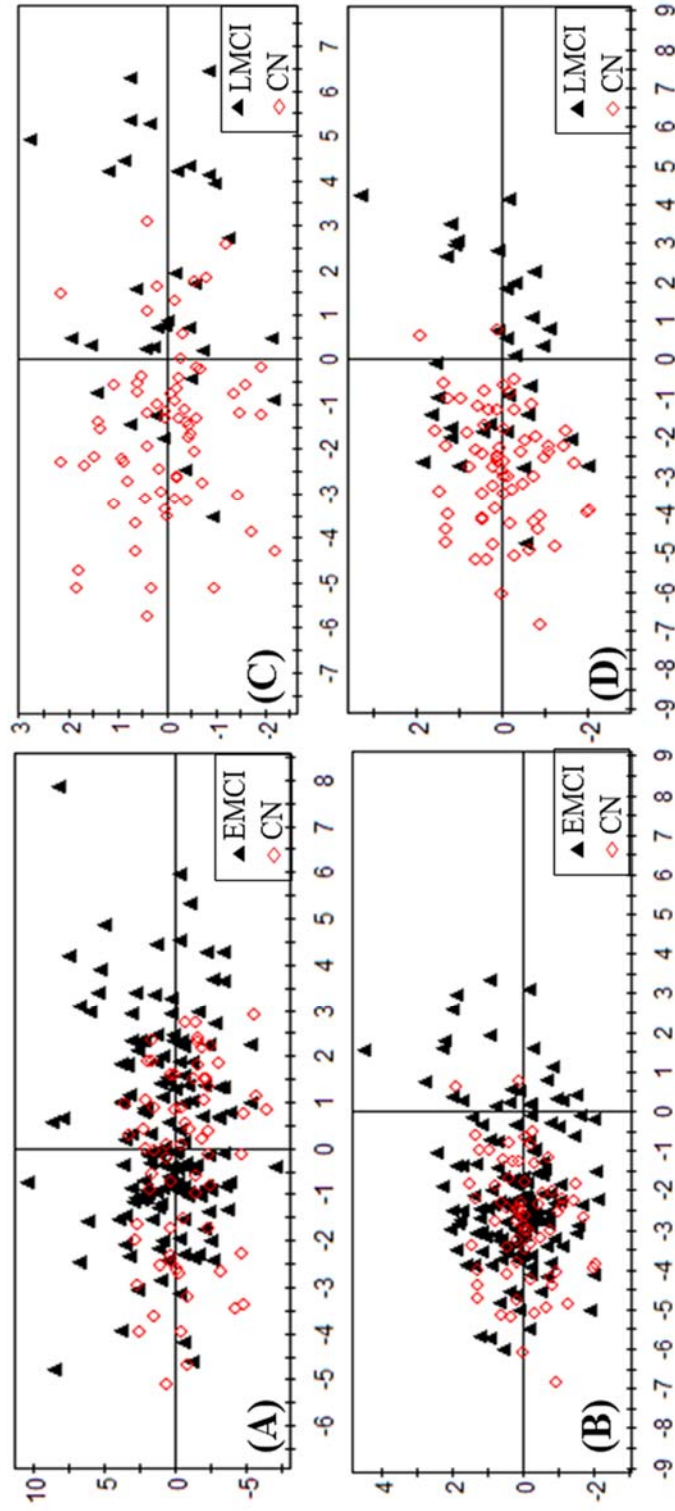


Figure 19: Scatter plots of best feature-selected models validated with external test set. (A) Testing best feature-selected EMCI model with external EMCI and CN subjects. (B) Testing best feature-selected AD model with external EMCI and CN subjects. (C) Testing best feature-selected LMCI model with external LMCI and CN subjects. (D) Testing best feature-selected AD model with external LMCI data and CN subjects.

Table 26: Classification performance (accuracy, sensitivity and specificity) for validation of best EMCI and LMCI models using external test set

	EMCI vs. CN			LMCI vs. CN		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
Best full model of EMCI	48.8%	48.6%	49.3%	-	-	-
Best feature-selected model of EMCI	49.3%	50.0%	47.8%	-	-	-
Best full model of LMCI	-	-	-	74.0%	69.7%	76.1%
Best feature-selected model of LMCI	-	-	-	83.0%	75.8%	86.6%
Best full AD-model	42.7%	18.1%	95.6%	80.0%	48.5%	95.6%
Best feature-selected model of AD	41.7%	16.0%	97.0%	82.0%	51.5%	97.0%

5.4 Discussion

5.4.1 Significance of MRI and PET features

This study considered MRI measures including hemispheric measures, various sorts of cortical and subcortical segmentations and parcellation (regional volumes, cortical thickness and surface area etc.), hippocampal subfields and select regional PET features (AV-45 PET and FDG PET), making direct comparison of variables in terms of their importance within and across imaging modalities.

The results in Table 20 to Table 24 confirmed with previous study that atrophy in temporal lobe was more pronounced than in other lobes and medial temporal atrophy was highly symmetric, which differ from other dementia, such as semantic dementia with symmetric atrophy (Chan, et al., 2001; Kohler, et al., 1998). In this study, the following

hemispheric regional measures were found to be significant for predicting all stages of the disease: cortical volumes of inferior temporal, cortical thickness of fusiform, temporal pole, subcortical volumes of hippocampus, inferior lateral ventricle and amygdala, hippocampal subfields of cornu ammonis 1 (CA 1) and CA 2-3, and PET feature of SUVR across the brain. This was consistent with our previous study that combining the first two of the aforementioned 3 subcortical volumes with the neuropsychological test score yielded an accuracy of 92.4% in predicting AD from CN(Zhou, et al., 2014d). Besides, discrepancy of regional significance among hippocampal subfields as demonstrated in Table 24 may have improved differential diagnosis of AD according to a study by Mueller and Weiner , who reported that subfield volumetry of hippocampus provides regional selective information for distinguishing different pathologies affecting the hippocampus (Mueller and Weiner, 2009). It is therefore helpful to include hippocampal subfields in multivariate study of AD.

Landau et al. found that beta-amyloid deposition has an early and subclinical impact on cognition preceding metabolic changes and that hypometabolism becomes more pronounced and more closely related to ongoing cognitive decline as disease progresses (Landau, et al., 2012). This is consistent with the results in Table 24 that summary of SUVR across the whole brain as defined earlier in the method section, was the only PET (AV-45 or FDG) feature showing to be significant when comparing any stages of AD with CN, and it is the only AV-45 PET feature that is significant in predicting EMCI. This may imply that at a very early stage of MCI, beta-amyloid deposition could be more pronounced brain-widely instead of regionally and at such a stage hypometabolism was

not prominent enough across the brain though showing certain impacts regionally. Wu et al. also reported similar results in that significant amyloid accumulation is presented in EMCI while brain metabolism remains normal (Wu, et al., 2012).

5.4.2 Model efficiency with OPLS

This study aimed to investigate the predictive power of MRI and select regional PET features in discriminating AD, LMCI, and EMCI from controls using OPLS as a multivariate analysis tool based on a prior OPLS study (Westman, et al., 2011b). Full models using all or one category of the features were created for EMCI, LMCI and AD against CN. Moreover, two feature selection techniques were implemented and feature-selected models were created. Cross-validated predictive power $Q_2(Y)$ was used to evaluate these models, which were also validated with external test set.

Efficiency of full models with OPLS: To the best of our knowledge about the literature, this study was the first to analyze EMCI and LMCI separately using OPLS as a multivariate tool, though study analyzing AD and MCI based on MRI measures using OPLS has been done. Using the European AddNeuroMed project data, Westman et al. utilized automated regional volumes and manual outlining of hippocampus as inputs to OPLS and found a $Q_2(Y)$ of 0.64 when discriminating AD from CN (Westman, et al., 2011b), which was consistent with the $Q_2(Y)$ of 0.645 achieved using only MRI in this study, the difference could be due to differences in MRI features used and population differences. In addition, we also showed that PET scans have comparable discriminative power ($Q_2(Y) = 0.636$) with MRI. Westman et al. also built a MCI model with the same

features and obtained a $Q_2(Y)$ of 0.22. Compared with LMCI model using all MRI features in this study, the predictive power of our model ($Q_2(Y) = 0.282$) was higher which we suspect to be mostly due to the difference between diagnostic criteria of MCI in AddNeuroMed dataset and LMCI in ADNI dataset, and could also be that MRI features included in this study contributed more complementary information to the model.

Interestingly, a model with cortical thickness measures was found to have the most discriminative power than any other MRI and PET features, even higher than combining all of them. The same case was not seen for LCMI and AD models, which showed that volumetric measures had more power. This could be due to the higher reliability of cortical thickness than of cortical volumes and other MRI measures at the very early stage of AD (Querbes, et al., 2009). This result is consistent with other studies that recommend using cortical thickness for MCI prediction (Querbes, et al., 2009; Wang, et al., 2009). This could also suggest that cortical thinning is the very first anatomical change that occurs before any other volumetric change or effective response to PET scan, and as disease progresses volumetric change will eventually dominate.

Since the discriminative power of the best EMCI model is still weak, it can be argued that this could be due to partitioning errors embedded in the cross validation process that the data distribution may slightly favor cortical thickness features than others as reflected in the $Q_2(Y)$ measure. However, $Q_2(Y)$ of 0.108 is higher than the threshold (0.05) for

significance of an OPLS model(Westman, et al., 2011b), therefore the model is still considered significant and the results are still considered reliable.

Efficiency of feature-selected models with OPLS: Feature selection was also investigated since it has been reported that feature selection improves classification performance if the right prior knowledge is used (Chu, et al., 2012). In this study, prior knowledge of VIP scores in PLS projection was used to iteratively exclude less significant features based on VIP scores to investigate the predictability of feature-selected models in comparison to full models which include all features. Feature selection based on p values was also of interest as it's a commonly used feature selection technique (Chaves, et al., 2009b; Chu, et al., 2012).

The results as shown in Figure 17 not only confirmed the results of a previous study (Chu, et al., 2012) in the use of feature selection based on prior knowledge, it also showed that recursive feature selection based on VIP scores was effective in enhancing classification performance better than just using Student's t test as the only selection criteria. The reason that models with select features did a better job than models using all features could be that some of the less significant features contribute little or even negatively to the separation given their inconsistent variations. It could also mean that large numbers of input variables may actually overfit the data (Chu, et al., 2012; Guyon and Elisseeff, 2003).

5.4.3 Model validation with external test set

In order to avoid potential bias due to sample size differences, the 4 compared groups in the training process were assigned the same sample size. As the AD group has the least amount of samples, testing group was not assigned any AD subjects in order to maximize statistical power by ensuring that sample size is as large as possible for the 4 groups in the training process. Thus the study should not be criticized for that, as such data distribution is to fully maximize validity of the results. In addition, the main focus of this study was to analyze the prodromal stages (EMCI and LMCI) of AD separately. To a certain extent, the results of this study could also partially imply robustness of AD models from the following aspects: firstly best full AD model demonstrated perfect separation of AD and CN with no errors in the training process and Table 26 showed that best AD models have indeed high specificities of more than 95% during validation using external test set.

Even though not reflected by the results of EMCI model validation in Table 26, the EMCI model displayed some generalization as shown in Figure 18A and Figure 19A that those residing far from the separation boundary on the EMCI side were mostly EMCI subjects, indicating that these EMCI subjects in the test set share similar pattern as EMCI subjects in the training set. This could also be an indicator of structural change of brain (cortical thickness in particular) already showing up at this stage. Overall, due to more consistency of LMCI patterns, LMCI models were more robust and had better generalization to external test set than EMCI models.

Combining AV-45 and FDG PET scans received a predictive power ($Q_2(Y) = 0.093$) close to that of the cortical thickness, indicating that amyloid deposition and hypometabolism have emerged at the EMCI stage. As a stage close to AD dementia, LMCI has its own pattern of atrophy that differentiates it from AD, which could be seen from Table 26 that best feature-selected model of LMCI predicted external LMCI subjects better than the best AD models. Table 26 also shows that best AD models are not as efficient as the best EMCI models in identifying EMCI subjects neither. The cause of the outcome is mainly due to the fact that significance of ROIs changes as disease progresses.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Conclusions from Analyzing Methods

6.1.1 Combining MRI with Neuropsychological Test (MMSE)

The use of different biomarkers for clinical diagnosis of AD and MCI is of great importance. This study has shown that volumetric MRI measures can better predict AD (aMCI or naMCI) when combined with MMSE score. The MMSE score is found to be the most statistically significant variable and one that improves classification accuracy at any stage of the AD spectrum by over 10%. Particularly MRI measures of right hippocampus and left inferior lateral ventricle when combined with MMSE score yield a classification accuracy of 92.4% (sensitivity: 84.0%; specificity: 96.1%) for delineating AD patients from CN, which is very competitive in comparison to results reported in other recent studies. The approach considered for selecting and then ranking MMSE and other MRI variables could be useful at augmenting other classification methods reported in the literature and could have broader impact in reevaluating the different variables as predictive measures of AD. In addition, the results show that AD atrophy is widely spread and evenly seen on both sides, whereas aMCI and naMCI subjects are left side and right side dominant, respectively, which indicates that at different stages of AD, there may be hemisphere-dependent atrophy dominance. Also hippocampus atrophy is found in this study to be one of the key factors causing memory problems among subjects with

aMCI as compared to naMCI, which builds a linkage between the distinctive symptoms of the two types of MCI and the brain atrophy, and strongly suggests the use of hippocampus atrophy as the means to separate aMCI from naMCI.

6.1.2 Investigating Normalization Effect

This dissertation also studied the effect of normalization on the proposed statistical feature selection approach using ROIs segmented by Freesurfer and a neuropsychological test in terms of classification performance. The results show that subcortical volume should not be normalized and surface area does not bear much discriminative information as compared to subcortical volumes or cortical thickness. Also, subcortical volumes and cortical thickness based brain maps of significant regions show symmetry between the two hemispheres which is not seen in the brain maps generated using surface area. Moreover, the feature selection method implemented on cortical thickness measures show that normalization to either ICV or mean thickness exhibits an enhancement on the classification performance, and the most pronounced changes in the cortical thickness related to AD is seen in the temporal lobe of the brain, which is shown to be related to symptoms in AD patients regarding organization, language, understanding, etc. A comparison of results using the optimal model which combines MMSE with subcortical volumes shows that the proposed study achieved a competitive accuracy of 92.3% using fewer biomarkers, which makes it cost-effective and convenient.

6.1.3 Combining MRI with PET

Based on prior study by Westman (Westman, et al., 2011b), this study utilized 271 MRI features and 15 pre-identified PET (AV-45 and FDG) features as inputs to a multivariate analysis tool (OPLS) to discriminate EMCI, LMCI and AD from controls based on ADNI database. The results showed that MRI and PET had similar predictive power of AD and cortical thickness is a significant measure in identifying EMCI subjects. Feature selection techniques were implemented to explore models with optimal discriminative power. The results demonstrated that OPLS is a powerful tool identifying AD and LMCI with great generalization. Due to insufficiency of consistent pattern between EMCI subjects and controls, EMCI prediction appeared to suffer from poor generalization. Recursive feature selection based on VIP enhanced prediction when features were properly selected. Two to three iterations of VIP-based feature selection were suggested for prediction of AD and LMCI, and a single iteration was sufficient for prediction of EMCI to achieve optimal predictability. In addition, models well trained with AD pattern were neither optimized for EMCI nor for LMCI predictions.

Such study of identifying EMCI and LMCI as independent stages has rarely been done in the literature as far as we know. As a result, heterogeneity between them has not yet fully been explored. Multivariate analysis of EMCI and LMCI as was done in this study could serve as an early step in seeking a solution to this problem, just as pursuing research for better biomarkers must go on. Since baseline image scans fail to properly predict subjects at EMCI stage using OPLS, future work to augment early findings of this study needs to

explore the progressive change of biomarkers (atrophy, amyloid deposition and hypometabolism level, among others) based on longitudinal scans using OPLS.

Future investigation on this research should be extended to utilize more valid biomarkers for AD, including PET, CSF, APOE and EEG.. Even though there are not enough subjects with all these biomarkers available, it would be essential to explore their complementary discriminative information to achieve better results and find connections between the different biomarkers. Another direction for future work would be to exhaustively explore machine learning algorithms for prediction, including feature selection methods (e.g. subset selection and using regularization) and classification methods (e.g. random forests and artificial neuron networks).

6.1.4 EMI Approach

There already existing a number of techniques that have been applied to MRI scans for AD prediction. However, they are mostly based on the techniques such to investigate the variance in the testing data explained by the variance found through training data as PCA and PLS etc. The prediction is then obtained using classifiers such as SVM. In this study, a novel EMI-based approach is proposed, which makes prediction based on a calculated value without data projection and such a value is directly related to the level of atrophy, which may potentially serve as an indicator of the severity of the disease. The results of this study demonstrated that the EMI-based method is very efficient and robust to external data for disease prediction using only MRI, outperforming even other prior studies that combined multiple biomarkers. The simplicity in implementation and the

high and consistent prediction performance suggest that the EMI-based method could potentially serve as cost-effective tool for the diagnosis of AD.

One important research direction for the future specifically for the EMI method is to develop a weighted averaging method for calculating EMI value. Since the higher ranked variables may contain more useful information in calculating the EMI and thus should be assigned a higher weighting factor. This procedure will be helpful in improving the classification performance but is also very challenging as different weighting methods may yield highly varying results and it is hard to justify which one performs better.

REFERENCES

- Achterberg, H.C., van der Lijn, F., den Heijer, T., Vernooij, M.W., Ikram, M.A., Niessen, W.J., de Bruijne, M. (2014) Hippocampal shape is predictive for the development of dementia in a normal, elderly population. *Human brain mapping*, 35:2359-2371.
- Apostolova, L.G., Green, A.E., Babakchian, S., Hwang, K.S., Chou, Y.-Y., Toga, A.W., Thompson, P.M. (2012) Hippocampal atrophy and ventricular enlargement in normal aging, mild cognitive impairment and Alzheimer's disease. *Alzheimer disease and associated disorders*, 26:17.
- Association, A.P., DSM-IV., A.P.A.T.F.o. (1994) Diagnostic and statistical manual of mental disorders: DSM-IV. Amer Psychiatric Pub Inc.
- Barnes, J., Ridgway, G.R., Bartlett, J., Henley, S.M., Lehmann, M., Hobbs, N., Clarkson, M.J., MacManus, D.G., Ourselin, S., Fox, N.C. (2010) Head size, age and gender adjustment in MRI studies: a necessary nuisance? *Neuroimage*, 53:1244-55.
- Besthorn, C., Förstl, H., Geiger-Kabisch, C., Sattel, H., Gasser, T., Schreiter-Gasser, U. (1994) EEG coherence in Alzheimer disease. *Electroencephalography and clinical neurophysiology*, 90:242-245.
- Bickeböller, H., Campion, D., Brice, A., Amouyel, P., Hannequin, D., Didierjean, O., Penet, C., Martin, C., Perez-Tur, J., Michon, A. (1997) Apolipoprotein E and Alzheimer disease: genotype-specific risks by age and sex. *American journal of human genetics*, 60:439.
- Blennow, K., Hampel, H. (2003) CSF markers for incipient Alzheimer's disease. *The Lancet Neurology*, 2:605-613.
- Bylesjo, M., Rantalainen, M., Cloarec, O., Nicholson, J.K., Holmes, E., Trygg, J. (2006) OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J Chemometr*, 20:341-351.
- Carne, R.P., Vogrin, S., Litewka, L., Cook, M.J. (2006) Cerebral cortex: An MRI-based study of volume and variance with age and sex. *J Clin Neurosci*, 13:60-72.

- Chan, D., Fox, N., Scahill, R., Crum, W., Whitwell, J., Ciolotti, L., Rossor, M.N. (2001) Patterns of temporal lobe atrophy in semantic dementia and Alzheimer's disease. *J Neurol Neurosurg Ps*, 70:276-276.
- Chaves, R., Ramírez, J., Górriz, J., López, M., Salas-Gonzalez, D., Álvarez, I., Segovia, F. (2009a) SVM-based computer-aided diagnosis of the Alzheimer's disease using t -test NMSE feature selection with feature correlation weighting. *Neurosci Lett*, 461:293-297.
- Chaves, R., Ramirez, J., Gorriz, J.M., Lopez, M., Salas-Gonzalez, D., Alvarez, I., Segovia, F. (2009b) SVM-based computer-aided diagnosis of the Alzheimer's disease using t -test NMSE feature selection with feature correlation weighting. *Neurosci Lett*, 461:293-297.
- Chiang, G.C., Insel, P.S., Tosun, D., Schuff, N., Truran-Sacrey, D., Raptentsetsang, S., Jack, C.R., Weiner, M.W., Neuroimaging, A.D. (2011) Identifying Cognitively Healthy Elderly Individuals with Subsequent Memory Decline by Using Automated MR Temporoparietal Volumes. *Radiology*, 259:844-851.
- Choi, S.R., Golding, G., Zhuang, Z., Zhang, W., Lim, N., Hefti, F., Benedum, T.E., Kilbourn, M.R., Skovronsky, D., Kung, H.F. (2009) Preclinical properties of 18F-AV-45: a PET agent for A β plaques in the brain. *Journal of Nuclear Medicine*, 50:1887-1894.
- Chu, C., Hsu, A.-L., Chou, K.-H., Bandettini, P., Lin, C. (2012) Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. *Neuroimage*, 60:59-70.
- Csernansky, J., Wang, L., Swank, J., Miller, J., Gado, M., McKeel, D., Miller, M., Morris, J. (2005) Preclinical detection of Alzheimer's disease: hippocampal shape and volume predict dementia onset in the elderly. *Neuroimage*, 25:783-792.
- Cuenod, C.A., Denys, A., Michot, J.L., Jehenson, P., Forette, F., Kaplan, D., Syrota, A., Boller, F. (1993) Amygdala atrophy in Alzheimer's disease. An in vivo magnetic resonance imaging study. *Arch Neurol*, 50:941-5.
- Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehericy, S., Habert, M.O., Chupin, M., Benali, H., Colliot, O., Neuroimaging, A.D. (2011) Automatic classification of

- patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. *Neuroimage*, 56:766-781.
- Dale, A.M., Fischl, B., Sereno, M.I. (1999a) Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage*, 9:179-94.
- Dale, A.M., Fischl, B., Sereno, M.I. (1999b) Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage*, 9:179-194.
- Davatzikos, C., Bhatt, P., Shaw, L.M., Batmanghelich, K.N., Trojanowski, J.Q. (2011) Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiology of aging*, 32:2322. e19-2322. e27.
- de Leon, M.J., George, A.E., Reisberg, B., Ferris, S.H., Kluger, A., Stylopoulos, L.A., Miller, J.D., La Regina, M.E., Chen, C., Cohen, J. (1989a) Alzheimer's disease: longitudinal CT studies of ventricular change. *AJR Am J Roentgenol*, 152:1257-62.
- de Leon, M.J., George, A.E., Stylopoulos, L.A., Smith, G., Miller, D.C. (1989b) Early marker for Alzheimer's disease: the atrophic hippocampus. *Lancet*, 2:672-3.
- Dickerson, B.C., Bakkour, A., Salat, D.H., Feczko, E., Pacheco, J., Greve, D.N., Grodstein, F., Wright, C.I., Blacker, D., Rosas, H.D. (2009a) The cortical signature of Alzheimer's disease: regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. *Cerebral cortex*, 19:497-510.
- Dickerson, B.C., Feczko, E., Augustinack, J.C., Pacheco, J., Morris, J.C., Fischl, B., Buckner, R.L. (2009b) Differential effects of aging and Alzheimer's disease on medial temporal lobe cortical thickness and surface area. *Neurobiol Aging*, 30:432-40.
- Du, A.T., Schuff, N., Amend, D., Laakso, M.P., Hsu, Y.Y., Jagust, W.J., Yaffe, K., Kramer, J.H., Reed, B., Norman, D., Chui, H.C., Weiner, M.W. (2001) Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and Alzheimer's disease. *J Neurol Neurosurg Ps*, 71:441-447.
- Duara, R., Loewenstein, D.A., Greig, M., Acevedo, A., Potter, E., Appel, J., Raj, A., Schinka, J., Schofield, E., Barker, W. (2010) Reliability and Validity of an Algorithm for the Diagnosis of Normal Cognition, MCI and Dementia: Implications for Multi-

- Center Research Studies. The American journal of geriatric psychiatry: official journal of the American Association for Geriatric Psychiatry, 18:363.
- Duchesne, S., Caroli, A., Geroldi, C., Barillot, C., Frisoni, G.B., Collins, D.L. (2008) MRI-based automated computer classification of probable AD versus normal controls. IEEE Trans Med Imaging, 27:509-20.
- Ellison, J.M., Harper, D.G., Berlow, Y., Zeranski, L. (2008) Beyond the "C" on MCI: Noncognitive symptoms in amnesic and non-amnesic mild cognitive impairment. Cns Spectrums, 13:66-72.
- Eriksson, L., Hermens, J.L., Johansson, E., Verhaar, H.J., Wold, S. (1995) Multivariate analysis of aquatic toxicity data with PLS. Aquatic Sciences, 57:217-241.
- Eriksson, L., Johansson, E., Kettaneh-Wold, N., Trygg, J., Wikström, C., Wold, S. (2006) Multi-and Megavariate Data Analysis: Part II: Advanced Applications and Method Extensions.
- Eskildsen, S.F., Coupé, P., García-Lorenzo, D., Fonov, V., Pruessner, J.C., Collins, D.L. (2013) Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. Neuroimage, 65:511-521.
- Ewers, M., Walsh, C., Trojanowski, J.Q., Shaw, L.M., Petersen, R.C., Jack, C.R., Feldman, H.H., Bokde, A.L.W., Alexander, G.E., Scheltens, P., Vellas, B., Dubois, B., Weiner, M., Hampel, H., Neuroimaging, N.A.A.s.D. (2012) Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. Neurobiol Aging, 33:1203-+.
- Fan, Y., Resnick, S.M., Wu, X., Davatzikos, C. (2008a) Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study. Neuroimage, 41:277-85.
- Fan, Y., Resnick, S.M., Wu, X.Y., Davatzikos, C. (2008b) Structural and functional biomarkers of prodromal Alzheimer's disease: A high-dimensional pattern classification study. NeuroImage, 41:277-285.

- Fennema-Notestine, C., Hagler, D.J., McEvoy, L.K., Fleisher, A.S., Wu, E.H., Karow, D.S., Dale, A.M. (2009) Structural MRI biomarkers for preclinical and mild Alzheimer's disease. *Human brain mapping*, 30:3238-3253.
- Ferman, T.J., Smith, G.E., Kantarci, K., Boeve, B.F., Pankratz, V.S., Dickson, D.W., Graff-Radford, N.R., Wszolek, Z., Van Gerpen, J., Uitti, R., Pedraza, O., Murray, M.E., Aakre, J., Parisi, J., Knopman, D.S., Petersen, R.C. (2013) Nonamnesic mild cognitive impairment progresses to dementia with Lewy bodies. *Neurology*.
- Fischl, B., Dale, A.M. (2000) Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc Natl Acad Sci U S A*, 97:11050-5.
- Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S. (2002a) Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33:341-355.
- Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M. (2002b) Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33:341-55.
- Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D. (2004a) Automatically parcellating the human cerebral cortex. *Cerebral cortex*, 14:11-22.
- Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A.M. (2004b) Automatically parcellating the human cerebral cortex. *Cereb Cortex*, 14:11-22.
- Folstein, M.F. (1990) The Birth of the Mms - a Citation Classic Commentary on Mini-Mental State - a Practical Method for Grading the Cognitive State of Patients for the Clinician by Folstein,M.F., Folstein,S.E., and Mchugh,P.R. *Cc/Clin Med*:14-14.
- Folstein, M.F., Folstein, S.E., McHugh, P.R. (1975) Mini-Mental State: a practical method for grading the cognitive state of patients for the clinician. Pergamon Press.

- Fox, N., Scahill, R., Crum, W., Rossor, M. (1999) Correlation between rates of brain atrophy and cognitive decline in AD. *Neurology*, 52:1687-1687.
- Fox, N.C., Freeborough, P.A., Rossor, M.N. (1996) Visualisation and quantification of rates of atrophy in Alzheimer's disease. *Lancet*, 348:94-97.
- Fox, N.C., Schott, J.M. (2004) Imaging cerebral atrophy: normal ageing to Alzheimer's disease. *Lancet*, 363:392-4.
- Fuld, P.A. (1982) Fuld object memory evaluation. Stoelting.
- Galindo-Prieto, B., Eriksson, L., Trygg, J. (2014) Variable influence on projection (VIP) for orthogonal projections to latent structures (OPLS). *J Chemometr.*
- Ganzer, S., Arlt, S., Schoder, V., Buhmann, C., Mandelkow, E.-M., Finckh, U., Beisiegel, U., Naber, D., Müller-Thomsen, T. (2003) CSF-tau, CSF-A β 1-42, ApoE-genotype and clinical parameters in the diagnosis of Alzheimer's disease: combination of CSF-tau and MMSE yields highest sensitivity and specificity. *Journal of neural transmission*, 110:1149-1160.
- Gomar, J.J., Bobes-Bascaran, M.T., Conejero-Goldberg, C., Davies, P., Goldberg, T.E., Neuroimaging, A.s.D. (2011) Utility of Combinations of Biomarkers, Cognitive Markers, and Risk Factors to Predict Conversion From Mild Cognitive Impairment to Alzheimer Disease in Patients in the Alzheimer's Disease Neuroimaging Initiative. *Arch Gen Psychiat*, 68:961-969.
- Gour, N., Felician, O., Didic, M., Koric, L., Gueriot, C., Chanoine, V., Confort-Gouny, S., Guye, M., Ceccaldi, M., Ranjeva, J.P. (2013) Functional connectivity changes differ in early and late-onset alzheimer's disease. *Human brain mapping*.
- Guyon, I., Elisseeff, A. (2003) An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3:1157-1182.
- Haase, L., Wang, M., Green, E., Murphy, C. (2013) Functional connectivity during recognition memory in individuals genetically at risk for Alzheimer's disease. *Human brain mapping*, 34:530-542.

- Han, X., Jovicich, J., Salat, D., van der Kouwe, A., Quinn, B., Czanner, S., Busa, E., Pacheco, J., Albert, M., Killiany, R. (2006) Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. *Neuroimage*, 32:180-194.
- Hansson, O., Zetterberg, H., Buchhave, P., Londos, E., Blennow, K., Minthon, L. (2006) Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitive impairment: a follow-up study. *The Lancet Neurology*, 5:228-234.
- Harasty, J., Double, K.L., Halliday, G.M., Kril, J.J., McRitchie, D.A. (1997) Language-associated cortical regions are proportionally larger in the female brain. *Arch Neurol-Chicago*, 54:171-176.
- Hardy, J., Allsop, D. (1991a) Amyloid deposition as the central event in the aetiology of Alzheimer's disease. *Trends in pharmacological sciences*, 12:383-388.
- Hardy, J., Allsop, D. (1991b) Amyloid Deposition as the Central Event in the Etiology of Alzheimers-Disease. *Trends Pharmacol Sci*, 12:383-388.
- Herholz, K., Salmon, E., Perani, D., Baron, J., Holthoff, V., Frölich, L., Schönknecht, P., Ito, K., Mielke, R., Kalbe, E. (2002) Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG PET. *Neuroimage*, 17:302-316.
- Higdon, R., Foster, N.L., Koeppe, R.A., DeCarli, C.S., Jagust, W.J., Clark, C.M., Barbas, N.R., Arnold, S.E., Turner, R.S., Heidebrink, J.L. (2004) A comparison of classification methods for differentiating fronto-temporal dementia from Alzheimer's disease using FDG-PET imaging. *Statistics in medicine*, 23:315-326.
- Hinrichs, C., Singh, V., Xu, G.F., Johnson, S.C., Neuroimaging, A.D. (2011) Predictive markers for AD in a multi-modality framework: An analysis of MCI progression in the ADNI population. *Neuroimage*, 55:574-589.
- Hirao, K., Ohnishi, T., Hirata, Y., Yamashita, F., Mori, T., Moriguchi, Y., Matsuda, H., Nemoto, K., Imabayashi, E., Yamada, M. (2005) The prediction of rapid conversion to Alzheimer's disease in mild cognitive impairment using regional cerebral blood flow SPECT. *Neuroimage*, 28:1014-1021.

- Hsiao, T., Huang, C.-C., Hsieh, C.-J., Hsu, W.-C., Wey, S.-P., Yen, T.-C., Kung, M.-P., Lin, K.-J. (2012) Correlation of early-phase 18F-florbetapir (AV-45/Amyvid) PET images to FDG images: preliminary studies. *European journal of nuclear medicine and molecular imaging*, 39:613-620.
- Huang, C., Wahlund, L., Dierks, T., Julin, P., Winblad, B., Jelic, V. (2000) Discrimination of Alzheimer's disease and mild cognitive impairment by equivalent EEG sources: a cross-sectional and longitudinal study. *Clin Neurophysiol*, 111:1961-7.
- Jack, C.R., Petersen, R.C., Xu, Y.C., Waring, S.C., O'Brien, P.C., Tangalos, E.G., Smith, G.E., Ivnik, R.J., Kokmen, E. (1997) Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. *Neurology*, 49:786-794.
- Jagust, W., Thisted, R., Devous, M., Van Heertum, R., Mayberg, H., Jobst, K., Smith, A., Borys, N. (2001) SPECT perfusion imaging in the diagnosis of Alzheimer's disease A clinical-pathologic study. *Neurology*, 56:950-956.
- Johnson, K.A., Jones, K., Holman, B., Becker, J., Spiers, P., Satlin, A., Albert, M. (1998) Preclinical prediction of Alzheimer's disease using SPECT. *Neurology*, 50:1563-1571.
- Jovicich, J., Czanner, S., Han, X., Salat, D., van der Kouwe, A., Quinn, B., Pacheco, J., Albert, M., Killiany, R., Blacker, D. (2009) MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. *Neuroimage*, 46:177-192.
- Juottonen, K., Laakso, M.P., Partanen, K., Soininen, H. (1999) Comparative MR analysis of the entorhinal cortex and hippocampus in diagnosing Alzheimer disease. *American Journal of Neuroradiology*, 20:139-144.
- Klöppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox, N.C., Jack, C.R., Ashburner, J., Frackowiak, R.S. (2008) Automatic classification of MR scans in Alzheimer's disease. *Brain*, 131:681-689.
- Kloppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox, N.C., Jack, C.R., Jr., Ashburner, J., Frackowiak, R.S. (2008) Automatic classification of MR scans in Alzheimer's disease. *Brain*, 131:681-9.

- Kohler, S., Black, S.E., Sinden, M., Szekely, C., Kidron, D., Parker, J.L., Foster, J.K., Moscovitch, M., Wincour, G., Szalai, J.P., Bronskill, M.J. (1998) Memory impairments associated with hippocampal versus parahippocampal-gyrus atrophy: an MR volumetry study in Alzheimer's disease. *Neuropsychologia*, 36:901-914.
- Laakso, M., Partanen, K., Riekkinen, P., Lehtovirta, M., Helkala, E.-L., Hallikainen, M., Hanninen, T., Vainio, P., Soininen, H. (1996) Hippocampal volumes in Alzheimer's disease, Parkinson's disease with and without dementia, and in vascular dementia An MRI study. *Neurology*, 46:678-681.
- Laakso, M., Soininen, H., Partanen, K., Helkala, E.-L., Hartikainen, P., Vainio, P., Hallikainen, M., Hänninen, T., Riekkinen Sr, P. (1995a) Volumes of hippocampus, amygdala and frontal lobes in the MRI-based diagnosis of early Alzheimer's disease: correlation with memory functions. *Journal of Neural Transmission-Parkinson's Disease and Dementia Section*, 9:73-86.
- Laakso, M.P., Soininen, H., Partanen, K., Helkala, E.L., Hartikainen, P., Vainio, P., Hallikainen, M., Hanninen, T., Riekkinen, P.J., Sr. (1995b) Volumes of hippocampus, amygdala and frontal lobes in the MRI-based diagnosis of early Alzheimer's disease: correlation with memory functions. *J Neural Transm Park Dis Dement Sect*, 9:73-86.
- Landau, S., Jagust, W. (2011a) Florbetapir processing methods.
- Landau, S., Jagust, W. (2011b) UC Berkeley FDG MetaROI methods.
- Landau, S.M., Harvey, D., Madison, C.M., Koeppe, R.A., Reiman, E.M., Foster, N.L., Weiner, M.W., Jagust, W.J., Initi, A.s.D.N. (2011) Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI. *Neurobiol Aging*, 32:1207-1218.
- Landau, S.M., Mintun, M.A., Joshi, A.D., Koeppe, R.A., Petersen, R.C., Aisen, P.S., Weiner, M.W., Jagust, W.J. (2012) Amyloid deposition, hypometabolism, and longitudinal cognitive decline. *Annals of neurology*, 72:578-586.
- Leclerc, B., Abulrob, A. (2013) Perspectives in Molecular Imaging Using Staging Biomarkers and Immunotherapies in Alzheimer's Disease. *Sci World J*.

- Locatelli, T., Cursi, M., Liberati, D., Franceschi, M., Comi, G. (1998) EEG coherence in Alzheimer's disease. *Electroencephalography and clinical neurophysiology*, 106:229-237.
- Loewenstein, D.A., Barker, W.W., Harwood, D.G., Luis, C., Acevedo, A., Rodriguez, I., Duara, R. (2000) Utility of a modified mini-mental state examination with extended delayed recall in screening for mild cognitive impairment and dementia among community dwelling elders. *Int J Geriatr Psych*, 15:434-440.
- López, M., Ramírez, J., Górriz, J., Álvarez, I., Salas-Gonzalez, D., Segovia, F., Chaves, R. (2009) SVM-based CAD system for early detection of the Alzheimer's disease using kernel PCA and LDA. *Neuroscience Letters*, 464:233-238.
- Lopez, M.M., Ramirez, J., Gorriz, J.M., Alvarez, I., Salas-Gonzalez, D., Segovia, F., Chaves, R. (2009) SVM-based CAD system for early detection of the Alzheimer's disease using kernel PCA and LDA. *Neurosci Lett*, 464:233-238.
- Luxenberg, J., Haxby, J.V., Creasey, H., Sundaram, M., Rapoport, S. (1987) Rate of ventricular enlargement in dementia of the Alzheimer type correlates with rate of neuropsychological deterioration. *Neurology*, 37:1135-1135.
- Magnin, B., Mesrob, L., Kinkingnehun, S., Pelegrini-Issac, M., Colliot, O., Sarazin, M., Dubois, B., Lehericy, S., Benali, H. (2009a) Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. *Neuroradiology*, 51:73-83.
- Magnin, B., Mesrob, L., Kinkingnehun, S., Pélégri-Issac, M., Colliot, O., Sarazin, M., Dubois, B., Lehericy, S., Benali, H. (2009b) Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. *Neuroradiology*, 51:73-83.
- Mattsson, N., Zetterberg, H., Hansson, O., Andreasen, N., Parnetti, L., Jonsson, M., Herukka, S.-K., van der Flier, W.M., Blankenstein, M.A., Ewers, M. (2009) CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment. *Jama*, 302:385-393.
- McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M. (1984) Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work

- Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, 34:939-44.
- Mosconi, L., Tsui, W.H., Herholz, K., Pupi, A., Drzezga, A., Lucignani, G., Reiman, E.M., Holthoff, V., Kalbe, E., Sorbi, S. (2008) Multicenter standardized 18F-FDG PET diagnosis of mild cognitive impairment, Alzheimer's disease, and other dementias. *Journal of Nuclear Medicine*, 49:390-398.
- Mudher, A., Lovestone, S. (2002a) Alzheimer's disease-do tauists and baptists finally shake hands? *Trends Neurosci*, 25:22-6.
- Mudher, A., Lovestone, S. (2002b) Alzheimer's disease-do tauists and baptists finally shake hands? *Trends in neurosciences*, 25:22-26.
- Mueller, S.G., Weiner, M.W. (2009) Selective effect of age, Apo e4, and Alzheimer's disease on hippocampal subfields. *Hippocampus*, 19:558-64.
- Nestor, S.M., Rupsingh, R., Borrie, M., Smith, M., Accomazzi, V., Wells, J.L., Fogarty, J., Bartha, R. (2008a) Ventricular enlargement as a possible measure of Alzheimer's disease progression validated using the Alzheimer's disease neuroimaging initiative database. *Brain*, 131:2443-2454.
- Nestor, S.M., Rupsingh, R., Borrie, M., Smith, M., Accomazzi, V., Wells, J.L., Fogarty, J., Bartha, R., Initi, A.s.D.N. (2008b) Ventricular enlargement as a possible measure of Alzheimers disease progression validated using the Alzheimers disease neuroimaging initiative database. *Brain*, 131:2443-2454.
- Nestor, S.M., Rupsingh, R., Borrie, M., Smith, M., Accomazzi, V., Wells, J.L., Fogarty, J., Bartha, R., Initiative, t.A.s.D.N. (2008c) Ventricular enlargement as a possible measure of Alzheimer's disease progression validated using the Alzheimer's disease neuroimaging initiative database. *Brain*, 131:2443-2454.
- Nobili, F., Salmaso, D., Morbelli, S., Girtler, N., Piccardo, A., Brugnolo, A., Dessi, B., Larsson, S.A., Rodriguez, G., Pagani, M. (2008) Principal component analysis of FDG PET in amnesic MCI. *European journal of nuclear medicine and molecular imaging*, 35:2191-2202.
- Nordberg, A., Rinne, J.O., Kadir, A., Langstrom, B. (2010a) The use of PET in Alzheimer disease. *Nat Rev Neurol*, 6:78-87.

- Nordberg, A., Rinne, J.O., Kadir, A., Långström, B. (2010b) The use of PET in Alzheimer disease. *Nat Rev Neurol*, 6:78-87.
- Nutter-Upham, K.E., Saykin, A.J., Rabin, L.A., Roth, R.M., Wishart, H.A., Pare, N., Flashman, L.A. (2008) Verbal fluency performance in amnesic MCI and older adults with cognitive complaints. *Arch Clin Neuropsychol*, 23:229-41.
- Organization, W.H. (2012) Dementia: Fact Sheet no. 362. Apr.
- Pennanen, C., Kivipelto, M., Tuomainen, S., Hartikainen, P., Hanninen, T., Laakso, M.P., Hallikainen, M., Vanhanen, M., Nissinen, A., Helkala, E.L., Vainio, P., Vanninen, R., Partanen, K., Soininen, H. (2004) Hippocampus and entorhinal cortex in mild cognitive impairment and early AD. *Neurobiol Aging*, 25:303-10.
- Piert, M., Koeppe, R.A., Giordani, B., Berent, S., Kuhl, D.E. (1996) Diminished glucose transport and phosphorylation in Alzheimer's disease determined by dynamic FDG-PET. *Journal of nuclear medicine: official publication, Society of Nuclear Medicine*, 37:201-208.
- Qin, J.B., Yang, T.B., Kong, F.J., Zhou, Q. (2013) Oral contraceptive use and uterine leiomyoma risk: a meta-analysis based on cohort and case-control studies. *Arch Gynecol Obstet*, 288:139-148.
- Querbes, O., Aubry, F., Pariente, J., Lotterie, J.-A., Démonet, J.-F., Duret, V., Puel, M., Berry, I., Fort, J.-C., Celsis, P. (2009) Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve. *Brain*, 132:2036-2047.
- Ramírez, J., Górriz, J., Segovia, F., Chaves, R., Salas-Gonzalez, D., López, M., Álvarez, I., Padilla, P. (2010) Computer aided diagnosis system for the Alzheimer's disease based on partial least squares and random forest SPECT image classification. *Neuroscience Letters*, 472:99-103.
- Ray, K.I. (2012) The Application of SIMCA P+ in Shotgun Metabolomics Analysis of ZIC® HILIC-MS Spectra of Human Urine-Experience with the Shimadzu IT-T of and Profiling Solutions Data Extraction Software. *Journal of Chromatography & Separation Techniques*.

- Ridha, B.H., Barnes, J., Bartlett, J.W., Godbolt, A., Pepple, T., Rossor, M.N., Fox, N.C. (2006) Tracking atrophy progression in familial Alzheimer's disease: a serial MRI study. *Lancet Neurol*, 5:828-34.
- Rombouts, S.A.R.B., Barkhof, F., Witter, M.P., Scheltens, P. (2000) Unbiased whole-brain analysis of gray matter loss in Alzheimer's disease. *Neurosci Lett*, 285:231-233.
- Scahill, R.I., Schott, J.M., Stevens, J.M., Rossor, M.N., Fox, N.C. (2002) Mapping the evolution of regional atrophy in Alzheimer's disease: unbiased analysis of fluid-registered serial MRI. *Proc Natl Acad Sci U S A*, 99:4703-7.
- Thompson, P.M., Hayashi, K.M., de Zubicaray, G.I., Janke, A.L., Rose, S.E., Semple, J., Hong, M.S., Herman, D.H., Gravano, D., Doddrell, D.M. (2004a) Mapping hippocampal and ventricular change in Alzheimer disease. *Neuroimage*, 22:1754-1766.
- Thompson, P.M., Hayashi, K.M., De Zubicaray, G.I., Janke, A.L., Rose, S.E., Semple, J., Hong, M.S., Herman, D.H., Gravano, D., Doddrell, D.M., Toga, A.W. (2004b) Mapping hippocampal and ventricular change in Alzheimer disease. *Neuroimage*, 22:1754-66.
- Trygg, J., Wold, S. (2002) Orthogonal projections to latent structures (O-PLS). *J Chemometr*, 16:119-128.
- van den Berg, R.A., Hoefsloot, H.C., Westerhuis, J.A., Smilde, A.K., van der Werf, M.J. (2006) Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC genomics*, 7:142.
- Walhovd, K., Fjell, A., Brewer, J., McEvoy, L., Fennema-Notestine, C., Hagler, D., Jennings, R., Karow, D., Dale, A. (2010a) Combining MR imaging, positron-emission tomography, and CSF biomarkers in the diagnosis and prognosis of Alzheimer disease. *American Journal of Neuroradiology*, 31:347-354.
- Walhovd, K.B., Fjell, A.M., Brewer, J., McEvoy, L.K., Fennema-Notestine, C., Hagler, D.J., Jennings, R.G., Karow, D., Dale, A.M., Initia, A.D.N. (2010b) Combining MR Imaging, Positron-Emission Tomography, and CSF Biomarkers in the Diagnosis and Prognosis of Alzheimer Disease. *Am J Neuroradiol*, 31:347-354.

- Walhovd, K.B., Fjell, A.M., Brewer, J., McEvoy, L.K., Fennema-Notestine, C., Hagler, D.J., Jr., Jennings, R.G., Karow, D., Dale, A.M. (2010c) Combining MR imaging, positron-emission tomography, and CSF biomarkers in the diagnosis and prognosis of Alzheimer disease. *AJNR Am J Neuroradiol*, 31:347-54.
- Wang, K., Liang, M., Wang, L., Tian, L., Zhang, X., Li, K., Jiang, T. (2007a) Altered functional connectivity in early Alzheimer's disease: A resting-state fMRI study. *Human brain mapping*, 28:967-978.
- Wang, L., Beg, F., Ratnanather, T., Ceritoglu, C., Younes, L., Morris, J.C., Csernansky, J.G., Miller, M.I. (2007b) Large deformation diffeomorphism and momentum based hippocampal shape discrimination in dementia of the Alzheimer type. *Medical Imaging, IEEE Transactions on*, 26:462-470.
- Wang, L., Goldstein, F.C., Veledar, E., Levey, A.I., Lah, J.J., Meltzer, C.C., Holder, C.A., Mao, H. (2009) Alterations in Cortical Thickness and White Matter Integrity in Mild Cognitive Impairment Measured by Whole-Brain Cortical Thickness Mapping and Diffusion Tensor Imaging. *Am J Neuroradiol*, 30:893-899.
- Wang, Z., Yan, C., Zhao, C., Qi, Z., Zhou, W., Lu, J., He, Y., Li, K. (2011) Spatial patterns of intrinsic brain activity in mild cognitive impairment and alzheimer's disease: A resting-state functional MRI study. *Human brain mapping*, 32:1720-1740.
- Weljie, A.M., Dowlatabadi, R., Miller, B.J., Vogel, H.J., Jirik, F.R. (2007) An inflammatory arthritis-associated metabolite biomarker pattern revealed by ¹H NMR spectroscopy. *Journal of proteome research*, 6:3456-3464.
- Westman, E., Aguilar, C., Muehlboeck, J.S., Simmons, A. (2013) Regional Magnetic Resonance Imaging Measures for Multivariate Analysis in Alzheimer's Disease and Mild Cognitive Impairment. *Brain Topogr*, 26:9-23.
- Westman, E., Muehlboeck, J., Simmons, A. (2012a) Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion. *Neuroimage*, 62:229-238.
- Westman, E., Muehlboeck, J.S., Simmons, A. (2012b) Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion. *Neuroimage*, 62:229-238.

- Westman, E., Muehlboeck, J.S., Simmons, A. (2012c) Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion. *Neuroimage*, 62:229-38.
- Westman, E., Simmons, A., Muehlboeck, J.S., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Soininen, H., Weiner, M.W., Lovestone, S., Spenger, C., Wahlund, L.O., Consortium, A., Initi, A.s.D.N. (2011a) AddNeuroMed and ADNI: Similar patterns of Alzheimer's atrophy and automated MRI classification accuracy in Europe and North America. *Neuroimage*, 58:818-828.
- Westman, E., Simmons, A., Zhang, Y., Muehlboeck, J., Tunnard, C., Liu, Y., Collins, L., Evans, A., Mecocci, P., Vellas, B. (2011b) Multivariate analysis of MRI data for Alzheimer's disease, mild cognitive impairment and healthy controls. *NeuroImage*, 54:1178-1187.
- Wong, D.F., Rosenberg, P.B., Zhou, Y., Kumar, A., Raymont, V., Ravert, H.T., Dannals, R.F., Nandi, A., Brašić, J.R., Ye, W. (2010) In vivo imaging of amyloid deposition in Alzheimer disease using the radioligand 18F-AV-45 (flobetapir F 18). *Journal of Nuclear Medicine*, 51:913-920.
- Wu, L.Y., Rowley, J., Mohades, S., Leuzy, A., Dauar, M.T., Shin, M., Fonov, V., Jia, J.P., Gauthier, S., Rosa-Neto, P., Initia, A.D.N. (2012) Dissociation between Brain Amyloid Deposition and Metabolism in Early Mild Cognitive Impairment. *Plos One*, 7.
- Zhang, D.Q., Wang, Y.P., Zhou, L.P., Yuan, H., Shen, D.G., Initia, A.D.N. (2011) Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage*, 55:856-867.
- Zhou, Q., Goryawala, M., Cabrerizo, M., Barker, W., Duara, R., Adjouadi, A. (2014a) Significance of Normalization on Anatomical MRI Measures in Predicting Alzheimer's Disease. *The Scientific World Journal*, 2014, Article ID 541802:1-12.
- Zhou, Q., Goryawala, M., Cabrerizo, M., Barker, W., Duara, R., Adjouadi, M. (2014b) Significance of normalization on anatomical MRI measures in predicting Alzheimer's disease. *ScientificWorldJournal*, 2014:541802.

Zhou, Q., Goryawala, M., Cabrerizo, M., Barker, W., Duara, R., Adjouadi, M. (2014c) Significance of Normalization on Anatomical MRI Measures in Predicting Alzheimer's Disease. The Scientific World Journal, 2014.

Zhou, Q., Goryawala, M., Cabrerizo, M., Wang, J., Barker, W., Loewenstein, D., Duara, R., Adjouadi, M. (2014d) An Optimal Decisional Space for the Classification of Alzheimer's Disease and Mild Cognitive Impairment.

VITA

QI ZHOU

- | | |
|------|---|
| 2011 | B.E., Electrical Engineering
Yanshan University
Qinhuangdao, China |
| 2015 | Ph.D. Candidate, Electrical Engineering
Florida International University
Miami, Florida |

PUBLICATIONS AND PRESENTATIONS

Zhou, Q., Goryawala, M., Cabrerizo, M., Wang, J., Barker, W., Duara, R., Adjouadi, M. (2014) An Optimal Decisional Space for the Classification of Alzheimer's Disease and Mild Cognitive Impairment. IEEE Transactions on Biomedical Engineering, vol.61, no.8, pp.2245,2253, Aug. 2014

Zhou, Q., Goryawala, M., Cabrerizo, M., Barker, W., Duara, R., Adjouadi, M. (2014) Significance of Normalization on Anatomical MRI Measures in Predicting Alzheimer's Disease," The Scientific World Journal, vol. 2014, Article ID 541802, 11 pages, 2014. doi:10.1155/2014/541802

Goryawala, M., Zhou, Q., Barker, W., Loewenstein, D., Duara R., Adjouadi, M. (2015) Inclusion of Neuropsychological Scores in Atrophy Models Improve Diagnostic Classification of Alzheimer's Disease and Mild Cognitive Impairment, Computational Intelligence and Neuroscience (Accepted), 2015

Goryawala, M., Duara, R., Loewenstein, D., Zhou, Q., Barker, W. and Adjouadi, M.(2015) Apolipoprotein-E4 (ApoE4) carriers show altered small-world properties in the default mode network of the brain, Biomedical Physics & Engineering Express (Accepted), 2015

Cabrerizo, M., Cabrera, A., Perez, J.O., de la Rua, J., Rojas, N., Zhou, Q., Pinzon-Ardila, A., Gonzalez-Arias S.M., Adjouadi, M. (2014) Induced Effects of Transcranial Magnetic Stimulation on the Autonomic Nervous System and the Cardiac Rhythm, The Scientific World Journal, vol. 2014, ID 349718, 2014.

Duara, R., Barker, W., Loewenstein, D., T. Greig, M., Rodriguez, R., Goryawala, M. Zhou, Q., Adjouadi M. (2015) Insights into cognitive aging and Alzheimer's disease using amyloid PET and structural MRI scans. Clinical and Translational Imaging, pp.1-10, Feb. 2015. doi:10.1007/s40336-015-0110-6

Sargolzaei, S., Cabrerizo, M., Goryawala, M., Chen, G., Zhou, Q., Duara, R., Barker, W., Adjouadi, M. (2014) A practical guideline for intracranial volume estimation in patients with Alzheimer's disease." BMC bioinformatics 16.Suppl 7:S8, 2015.

Zhou, Q., Goryawala, M., Cabrerizo, M., Wang, J., Barker, W., Duara, R., Adjouadi, M. (2013) Combining Anatomical Biomarkers with Neuropsychological Data for Multidimensional Classification of Alzheimer's Disease, International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV'13), vol. 1, pp. 255-261, Las Vegas, July, 2013.

Zhou, Q., Goryawala, M., Cabrerizo, M., Wang, J., Barker, W., Duara, R., Adjouadi, M. (2013) Regional MRI measures and neuropsychological test for multi-dimensional analysis in Alzheimer's disease, The 6th International IEEE/EMBS Conference on Neural Engineering (NER), pp. 533-536, San Diego, Nov. 2013.

Chaparro, G., Zhou, Q., Liu, C., Tang, J., Liu, S. (2012) Power-Efficient Schemes via Workload Characterization on the Intel's Single-Chip Cloud Computer, The IEEE 26th International Parallel and Distributed Processing Symposium Workshops & PhD Forum (IPDPSW) 2012, pp. 999-1006, Shanghai, January, 2012.

Zhou, Q., Goryawala, M., Cabrerizo, M., Wang, J., Barker, W., Duara, R., Adjouadi, M., Utility of Regional [18F] – Florbetapir PET Imaging SUVRs in Discriminating Alzheimer's Disease and its Prodromal Stages. The 8th Human Amyloid Imaging (HAI), vol. 1, pp 113, Miami Beach, January, 2014

Wang, J., Eddin, A.S., Zhou, Q., Gaillard, W., Adjouadi, M. (2013) Discriminating regional functional networks in pediatric epilepsy, The 6th International IEEE/EMBS Conference on Neural Engineering (NER), pp. 1513-1516, San Diego, November, 2013.

Zhou, Q., Goryawala, M., Cabrerizo, M., Barker, W., Loewenstein, D., Duara, R., Adjouadi, M., Multivariate Analysis of Structural MRI and PET (FDG and 18F-AV-45) for Alzheimer's Disease and Its Prodromal Stages, The 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Chicago, August, 2014.

Goryawala, M., Zhou, Q., Duara, R., Loewenstein, D., Cabrerizo, M., Barker, W., Adjouadi, M. Altered Small-World Anatomical Networks in Apolipoprotein-E4 (ApoE4) Carriers using MRI, The 36th Annual Int'l Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Chicago, August, 2014.

Li, C., Zhou, Q., Goryawala, M., Loewenstein, D., Barker, W., Duara R., Adjouadi, M.(2015) Investigating the Utility of FDG and AV45 PET in both Two-classification and Multi-classification of Alzheimer's Disease and Its Prodromal Stages, The 9th Human Amyloid Imaging (HAI), January, 2015.