Large-scale neuroimaging in Alzheimer’s disease and normal aging

Xinyang Feng

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2019

Abstract

Large-scale neuroimaging in Alzheimer's disease and normal aging

Xinyang Feng

Large-scale neuroimaging data is becoming increasingly available, providing a rich data source with which to study neurological conditions. In this thesis, I demonstrate the utility of large-scale neuroimaging as it applies to Alzheimer's disease (AD) and normal aging, using univariate parametric mapping, regional analysis, and advanced machine learning. Specifically, this thesis covers: 1) validation and extension of prior studies using large-scale datasets; 2) AD diagnosis and normal aging evaluation empowered by large-scale datasets and advanced deep learning algorithms; 3) enhancement of cerebral blood volume (CBV) fMRI utility with retrospective CBV-fMRI technique.

First, I demonstrated the utility of large-scale datasets for validating and extending prior studies using univariate analytics. I presented a study localizing AD-vulnerable regions more reliably and with better anatomical resolution using data from more than 350 subjects. Following a similar approach, I investigated the structural characteristics of healthy APOE ε4 homozygous subjects screened from a large-scale community-based study. To study the neuroimaging signatures of normal aging, we performed a large-scale joint CBV-fMRI and structural MRI study covering age 20-70s, and a structural MRI study of normal aging covering the full age-span, with the elder group screened from a large-scale clinic-based study ensuring no evidence of AD using both longitudinal follow-up and cerebrospinal fluid (CSF) biomarkers evidences.

Second, I performed deep learning neuroimaging studies for AD diagnosis and normal aging evaluation, and investigated the regionality associated with each task. I developed an AD diagnosis method using a 3D convolutional neural network model trained and evaluated on ~4,600 structural MRI scans and further investigated a series of novel regionality analyses. I further extensively studied the utility of the structural MRI summary measure derived from the deep learning model in prodromal AD detection. This study constitutes a general analytic framework, which was followed to evaluate normal aging by performing deep learning-based age estimation in cognitively normal population using more than 6,000 scans. The deep learning neuroimaging models classified AD and estimated age with high accuracy, and also revealed regional patterns conforming to neuropathophysiology. The deep learning derived MRI measure demonstrated potential clinical utility, outperforming other AD pathology measures and biomarkers. In addition, I explored the utility of deep learning on positron emission tomography (PET) data for AD diagnosis and regionality analyses, further demonstrating the broad utility and generalizability of the method.

Finally, I introduced a technique enabling CBV generation retrospectively from clinical contrast-enhanced scans. The derivation of meaningful functional measures from such clinical scans is only possible through calibration to a reference, which was built from the *largest* collection of research CBV-fMRI scans from our lab. This method was validated in an epilepsy study and demonstrated the potential to enhance the utility of CBV-fMRI by enriching the CBV-fMRI dataset. This technique is also applicable to AD and normal aging studies, and potentially enables deep learning based analytic approaches applied on CBV-fMRI with similar pipelines used in structural MRI.

Collectively, this thesis demonstrates how mechanistic and diagnostic information on brain disorders can be extracted from large-scale neuroimaging data, using both classical statistical methods and advanced machine learning.

Table of Contents

Table of Figures

Table of Tables

Acknowledgments

I acknowledge all the investigators and funding agencies who provide valuable local and external datasets that made this thesis possible.

Data used in preparation of this thesis were in part obtained from the Washington Heights-Inwood Columbia Aging Project (WHICAP). Data collection and sharing was supported by the WHICAP funded by the National Institute on Aging (NIA) (PO1AG07232, R01AG037212, RF1AG054023).

Data used in preparation of this thesis were in part obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). Data collection and sharing was funded by the ADNI (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Data used in the preparation of this thesis were in part obtained from the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL) funded by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) which was made available at the ADNI database (www.loni.usc.edu/ADNI).

Data used in preparation of this thesis were in part obtained from the Frontotemporal Lobar Degeneration Neuroimaging Initiative (FTLDNI) database (4rtni-ftldni.ini.usc.edu). Data collection and sharing was funded by the FTLDNI (National Institutes of Health Grant R01 AG032306). The study is coordinated through the University of California, San Francisco, Memory and Aging Center. FTLDNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Data used in the preparation of this thesis were in part obtained from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/data). For up-to-date information on the study, visit www.ppmi-info.org. PPMI - a public-private partnership is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, the full names of all of the PPMI funding partners can found at www.ppmi-info.org/fundingpartners.

Data used in the preparation of this thesis were in part provided by the Cambridge Centre for Ageing and Neuroscience (Cam-CAN). Cam-CAN funding was provided by the UK

Biotechnology and Biological Sciences Research Council (grant number BB/H008217/1), together with support from the UK Medical Research Council and University of Cambridge, UK.

Data were provided in part by IXI, accessed from http://brain-development.org/ixi-dataset/.

Data were provided in part by OASIS. OASIS Cross-Sectional: Principal Investigators: D. Marcus, R, Buckner, J, Csernansky J. Morris; P50 AG05681, P01 AG03991, P01 AG026276, R01 AG021910, P20 MH071616, U24 RR021382. OASIS: Longitudinal: Principal Investigators: D. Marcus, R, Buckner, J. Csernansky, J. Morris; P50 AG05681, P01 AG03991, P01 AG026276, R01 AG021910, P20 MH071616, U24 RR021382.

Data were provided in part by the Brain Genomics Superstruct Project of Harvard University and the Massachusetts General Hospital, (Principal Investigators: Randy Buckner, Joshua Roffman, and Jordan Smoller), with support from the Center for Brain Science Neuroinformatics Research Group, the Athinoula A. Martinos Center for Biomedical Imaging, and the Center for Human Genetic Research. 20 individual investigators at Harvard and MGH generously contributed data to the overall project.

Data used in preparation of this thesis were in part obtained from the SchizConnect database (schizconnect.org). Data collection and sharing was funded by NIMH cooperative agreement 1U01 MH097435.

Other datasets that contribute to this thesis work include Southwest University Adult life-span Dataset (SALD), Consortium for Reliability and Reproducibility (CoRR), Dallas life-span Brain Study (DLBS), Southwest University Longitudinal Imaging Multimodal Brain Data Repository (SLIM).

Preface

This dissertation is the result of numerous collaborative efforts with exceptional colleagues. I would like to take this chance to express my gratitude for all the helps I have received over the past five years, which are great gifts I will carry in my future life.

First and foremost, I would like to thank my advisor Dr. Scott Small for providing me with this valuable opportunity to pursue my interest in neuroimaging and machine learning research in my PhD study. Dr. Small has been an attentive mentor and a role model with his diligence, character, scientific insights and his perpetual momentum to learn and embrace new things.

I would like to thank my co-advisor Dr. Andrew Laine for his encouragement and invaluable advice over the years. I really appreciate the support when I first started the explorations of deep learning, and the opportunities participating into the engineering communities.

I am very thankful for my thesis proposal and defense committee members: Dr. Paul Sajda, Dr. Thomas Vaughan, Dr. Adam Brickman. Thanks to Dr. Sajda for being on my committee from qualification exam to thesis proposal and defense, and all the helpful suggestions and critical thinking along the way, in addition to what I have learned being a student and a teaching assistant for his classes. Dr. Vaughan's research of promoting MRI to the world is an encouragement and inspiration for my research into the widely-available structural MRI. Dr. Brickman has provided me with numerous thoughtful comments and insights, and we have been involved in many helpful discussions over the years.

I would like to specially thank two former BME PhD students in the lab and now Professors at Columbia: Dr. Frank Provenzano and Dr. Jia Guo, with whom I had countless interesting and thoughtful conversations from neuroscience, imaging, machine learning to every aspect of daily life. Dr. Provenzano developed solid work in CBV-fMRI processing. Dr. Guo developed a set of sophisticated analytic tools in mouse imaging that enabled our translational research.

I would like to extend my thanks to all the great Small lab members: Andrea Urban, Hannah Sigmon, Melanie Jalloh, Jeanelle France, Dr. Rakshathi Basavaraju, Dr. Diego Berman, Dr. Sabrina Simoes, Dr. Yasir Qureshi, Dr. Jessica Neufeld, who together have made PH-18-400 area a vibrant working place.

I would also specially thank Dr. Elsa Angelini, co-director of Heffner Biomedical Imaging Lab. Dr. Angelini has provided me with many methodological insights and concrete solutions.

I would like to extend my thanks to senior members from Heffner Biomedical Imaging Lab: Dr. Ming Jack Po, Dr. Viktor Gamarnik, Dr. Guillaume David, Dr. Arthur Mikhno, Dr. Jingkuan Song, for making ET-373 a geeky and vigorous base.

I am grateful for all my extraordinary collaborators: Dr. Marla Hamberger from Department of Neurology for the valuable data that helped validate our idea about retrospective CBV; Dr. Ray Razlighi from Department of Neurology for all the fMRI knowledge I have learnt in his class as a student and a teaching assistant; Dr. Jeffrey Lieberman, Dr. Richard Sloan, Dr. Melanie Wall, Dr. Ragy Girgis from New York State Psychiatric Institute for all the insightful meetings; Dr. Zachary Lipton from Carnegie Mellon University for sharing with me rich knowledge of deep learning and scientific writing.

I am indebted to my family, especially to my parents Guichen Feng and Ruimin Wang for the academic environment since childhood, and their wholehearted love and support. Most importantly, I am heartily grateful for my loving wife Jie Yang, for the seven years we have been advancing together, for the five years of PhD voyage sharing scenery and storm, for all your patience, support, inspiration and love, which together make me a better person. This thesis is dedicated to my family.

New York, NY, May 12th, 2019

Xinyang Feng

1 Introduction

1.1 Background

1.1.1 Neuroimaging

Neuroimaging, more specifically brain imaging, is the technique to image the structure and function of the brain. Neuroimaging serves as a critical tool in neurological clinical practice as well as neuroscience investigations. Some of the commonly used neuroimaging modalities are shown in Figure 1-1. In this thesis, I will mainly focus on two of the modalities: T1 weighted structural MRI and steady state exogenous contrast cerebral blood volume functional magnetic resonance imaging (CBV-fMRI).

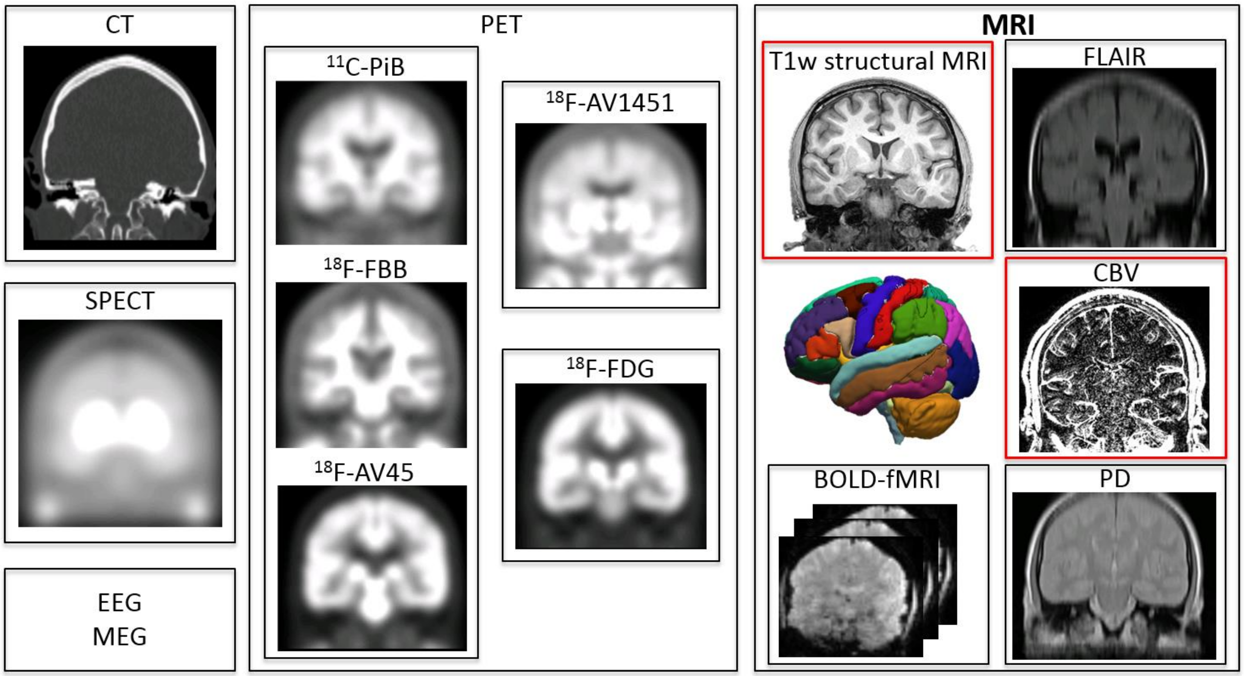


Figure 1-1. Sample neuroimaging modalities, with focuses on T1w structural MRI and CBV-fMRI in this thesis work.

1.1.2 Cerebral blood volume

Functional neuroimaging refers to studying the function and metabolism of the brain using imaging method, and broadly covers different variants of fMRI, positron emission tomography (PET), single-photon emission computed tomography (SPECT), electroencephalogram (EEG), magnetoencephalography (MEG), etc., among which fMRI is especially suitable to study brain regional functions because of its high spatial resolution and other practical aspects. FMRI can be roughly categorized into CBV, cerebral blood flow (CBF), Blood-oxygen-level dependent (BOLD) fMRI based on different underlying imaging measures.

Cerebral blood volume is a measure of the quantity of blood in a unit of tissue. A variant of MRI generated from steady-state exogenous contrast brain imaging using intravenous contrast agent (e.g. gadolinium-based contrast agents, GBCAs) can be used to generate CBV values on a voxel by voxel basis (Lin, Celik, & Paczynski, 1999). Values of percentage CBV extracted from these scans have been shown to reflect a basal state of cerebral activity, and CBV has been shown to strongly correlate with 18-flourodeoxyglucose PET (FDG-PET) measuring glucose metabolism (González et al., 1995).

CBV-fMRI enjoys the advantage of high spatial resolution, compared with other fMRI modalities, which is crucial to investigate complex neuroanatomical circuits such as the hippocampal subfields. Our lab has utilized CBV-fMRI to study the hippocampal dysfunctions in various brain disorders including Alzheimer’s disease (Khan et al., 2014), schizophrenia (Schobel et al., 2013), Huntington’s disease (Lewandowski et al., 2013), and also normal aging (Brickman et al., 2014).

1.1.3 Structural measurement of brain

Besides brain function and metabolism measured using CBV-fMRI and other functional neuroimaging modalities, the structural measurement of brain is also crucial to phenotype the brain pathophysiology, as it reflects the volume and composition of various cellular elements. Among different imaging modalities to quantify brain structures, I mainly focus on T1-weighted structural MRI, which is one of the most widely available and standardized imaging modalities and commonly has high spatial resolution and large contrast to noise ratio (CNR).

1.1.4 Alzheimer's disease and normal aging

Alzheimer's disease (AD) is a progressive neurodegenerative disease accounting for the majority cases of dementia (Alzheimer's Association, 2018), there are more than 50 million people living with dementia worldwide (Patterson, 2018). AD patients present progressive metabolic changes and atrophy in brain regions. Neurofibrillary tangles (NFTs) starting from entorhinal cortex (EC) or more focally transentorhinal cortex (TEC) characterize the early stage of AD according to postmortem studies (Braak & Braak, 1991). Amyloid, tau, and neurodegeneration are the major categories of AD biomarkers (Clifford R. Jack et al., 2018).

Normal aging in the context of this thesis refers to age-related changes in the brain that occur independent of any brain diseases. And the underlying pathophysiology is primarily dendritic spine loss without neuronal death (Morrison & Hof, 1997). Hippocampus is a subcortical structure heavily implicated in normal aging (Burke & Barnes, 2006; Scott A Small, 2011). Prefrontal cortex is another major region involved in normal aging (Gazzaley, Cooney, Rissman, & D'Esposito, 2005).

1.1.5 CBV-fMRI studies on regional vulnerability in Alzheimer's disease and normal aging

Prior studies in the lab have pinpointed transentorhinal cortex as the vulnerable region in Alzheimer's disease (Khan et al., 2014) and dentate gyrus (DG) as the vulnerable region in normal aging (Brickman et al., 2014) in the hippocampal formation using CBV-fMRI. Pioneering and impactful in our understanding of the underlying pathophysiology of Alzheimer's disease and normal aging, the findings from these prior studies could still be validated and extended using larger-scale datasets and different measurements such as structural measures.

1.1.6 Large-scale neuroimaging

There are increasing high-quality large-scale neuroimaging datasets with the advancement of MRI hardware, data storage, and the continuous efforts of numerous investigators. Some of the prominent projects with well-characterized participants include Alzheimer's Disease Neuroimaging Initiative (ADNI)[1], the Washington Heights-Hamilton Heights-Inwood Community Aging Project (WHICAP), etc. Over the years, our lab has collected the largest collection of research CBV-fMRI scans. And there are numerous large-scale studies focusing on other neurological diseases not listed here. These large-scale neuroimaging datasets greatly enrich the study population and cover more clinical and neurophysiological heterogeneity, allowing not only more reliable validation studies, but also more practical biomarker and diagnostics

---

[1] http://adni.loni.usc.edu/

development. It should be noted that, “large-scale” itself is a relative term that depends on the availability of the data modality, and is heuristically defined in (Fritsch et al., 2015) as more than 100.

1.1.7 Deep learning

Besides the interest of mechanistic understanding of brain disorders and normal aging, large-scale datasets also empower computer-aided imaging-biomarker discovery for diagnosis. Recent years have witnessed the rise of deep learning techniques (Goodfellow, Bengio, & Courville, 2016; LeCun, Bengio, & Hinton, 2015), the core of current artificial intelligence (AI) techniques. Deep learning, aka deep neural network, relies on the deep hierarchical representation enabled by efficient algorithm and hardware and large-scale datasets, and has revolutionized many aspects of research and everyday life including medical imaging (Gulshan et al., 2016; Hazlett et al., 2017). In the context of deep learning, “large-scale” manifests in both data and feature dimensions. Deep convolutional neural network (CNN) is the dominant class of deep learning model in image recognition field, and characterized by its efficiency at extracting patterns from images with weight-sharing and local connectivity properties, inspired by some early neuroscience studies in visual cortex (Hubel & Wiesel, 1962, 1968). And besides the superior discrimination performance, the interpretability of the machine learning models is receiving increasing attention (Lipton, 2018), especially in the medical diagnosis domain. It is possible, through interpretability analysis, the highly accurate machine learning model trained with broad population distribution can extract and summarize disease-defining patterns and inform potential disease pathophysiology.

1.2 Motivation

The advent of large-scale datasets, together with high-performance computing hardware and advanced machine learning algorithms, have opened up many new possibilities in neuroimaging research. In this thesis, I focus on the utilities of large-scale neuroimaging and demonstrate their feasibility and effectiveness as applied in Alzheimer's disease and normal aging studies: 1) validation and extension of prior studies using large-scale dataset; 2) neuroimaging based diagnosis and regionality analyses empowered by large-scale datasets and advanced deep learning algorithms; 3) enhancement of cerebral blood volume (CBV) fMRI utility with retrospective CBV.

1.3 Specific aims

More specifically, I target three specific aims in this thesis, which correspond to the three utilities of large-scale neuroimaging datasets listed in the motivation section.

Specific Aim 1 investigates large-scale validation and extension of prior studies using classical image processing and statistical methods, including 1) a large-scale structural neuroimaging study of AD that validates and localizes more reliably and with better anatomical resolution the AD-vulnerable region, and a structural neuroimaging study of healthy APOE ε4 homozygous subjects screened from a large-scale community-based study following a similar approach; 2) a large-scale joint functional and structural neuroimaging study of normal aging covering age 20-70s, and a structural neuroimaging study of normal aging covering the full age-span, with the elder group screened from a large-scale clinic-based study ensuring no evidence of AD using both longitudinal follow-up and CSF biomarkers evidences.

Specific Aim 2 focuses on deep learning neuroimaging. Without the time- and labor-consuming MRI post-processing and manual editing steps, the scale of the dataset can be further boosted. In this study, I proposed an AD diagnosis method using a 3D convolutional neural network model trained and evaluated on ~4,600 structural MRI scans and investigated a series of novel regionality analyses. This study constitutes a general analytic stream/framework, which was followed to perform a study on age estimation in normal aging population using more than 6,000 structural MRI scans. I demonstrated that a deep learning neuroimaging approach can achieve high classification and regression performance and also reveals patterns conforming to neuropathphysiological priors. The deep learning derived MRI measure demonstrated potential clinical utility, outperforming other AD pathology measures and biomarkers. In addition, I explored the utility of deep learning on PET data for AD diagnosis and regionality analyses, further demonstrating the broad utility and generalizability of the method.

Specific Aim 3 introduced retrospective CBV generation from clinical contrast-enhanced scans. The derivation of meaningful functional measures from such clinical scans is only possible through calibration to a reference, which was built from the collection of research CBV-fMRI scans from our lab. Validated in an epilepsy study, the method is also applicable to AD and normal aging studies. This technique is also applicable to AD and normal aging studies, and potentially enables deep learning based analytic approaches applied on CBV-fMRI with similar pipelines used in structural MRI.

1.4 Innovations

This thesis work has the potential to advance neuroimaging studies using large-scale datasets by demonstrating the feasibility and utility of extracting mechanistic and diagnostic information on brain disorders using both classical statistical and advanced machine learning methods. Large-scale studies could validate and extend the findings in prior studies in a more reliable and generalizable way. Large-scale datasets also enable studying low prevalence conditions and subjects screened with stringent criteria, through large well-characterized populations. Deep learning neuroimaging demonstrates superior performance in AD classification and age estimation, enjoys practical advantages, and is supported by the neuropathological priors, thus could potentially be developed into a practical multi-faceted neurological disease and normal aging evaluation system. The large-scale collection of CBV-fMRI scans makes it possible to mine underutilized clinical contrast-enhanced scans, and could greatly enhance the utility of CBV-fMRI, with the potential of applying advanced machine learning model in the future.

2 Specific Aim 1: Large-scale validation and extension of prior studies

In this section, I mostly relied on classical image processing and statistical methods to perform mechanistic studies of AD and normal aging with large-scale neuroimaging datasets.

I performed a study localizing AD-vulnerable regions more reliably and with better anatomical resolution using data from more than 350 subjects. Following a similar approach, I investigated the structural characteristics of cognitively normal APOE ε4 homozygous subjects screened from a large-scale community-based study.

To study the neuroimaging signatures of normal aging, we performed a large-scale joint CBV-fMRI and structural MRI study covering age 20-70s, and a structural MRI study of normal aging covering the full age-span, with the elder group screened from a large-scale clinic-based study ensuring no evidence of AD using both longitudinal follow-up and CSF biomarkers evidences.

2.1 Alzheimer's disease

2.1.1 Introduction

2.1.1.1 Transentorhinal cortex

Prior studies have pinpointed TEC as a vulnerable region in AD using CBV-fMRI, but because of the trade-off of CBV-fMRI in its high in-plane resolution (0.78 mm × 0.78 mm) and low slice thickness (3 mm), it would be less accurate when trying to interrogate the long axis, which also present specialization (Poppenk, Evensmoen, Moscovitch, & Nadel, 2013). We collected structural MRI scans with isotropic 1mm spatial resolution in a large-scale dataset comprised of

188 AD scans and 169 cognitively normal control scans from ADNI to localize the structurally AD-vulnerable region.

2.1.1.2 APOE

Apolipoprotein E (APOE) ε4 is one major genetic risk factor for AD. APOE ε4 homozygous subjects were shown to have medium-to-high risk of developing AD and low population prevalence (Scheltens et al., 2016). Studying the brain morphology of cognitively normal subjects in this high-risk group is important to understand the mechanism of the increased risk, disentangling from the effect of the developed AD pathology. Few studies investigate the cognitively normal elder APOE ε4 *homozygous* subjects because of the low prevalence. A number of previous studies on cognitively normal APOE ε4 carriers have mainly relied on middle-aged (Cacciaglia et al., 2018; Fennema-Notestine et al., 2011) and adolescent cohort (P. Shaw et al., 2007) to exclude the confound of AD, but age is arguably another risk factor that might interact with genetic risk factors.

2.1.2 Methods

2.1.2.1 ADNI dataset

The demographic information of the subjects used in the cortical thickness analysis to localize AD-vulnerable region is shown in Table 2-1. The details of the ADNI MRI acquisition protocols (Clifford R. Jack et al., 2008) can be found in the project website[2]. The T1w structural images

---

[2] http://adni.loni.usc.edu/adni-mri-methods/

were processed using FreeSurfer, generating individual cortical thickness maps (Bruce Fischl & Dale, 2000). The individual surfaces were registered into the FreeSurfer *fsaverage* space for inter-subject analysis. At each vertex, I performed linear regression with thickness as the dependent variable, AD vs. cognitively normal diagnosis as testing variable, with gender and age as covariates. The t-values of the regression analyses were rendered as t-maps in difference spaces.

Table 2-1. The demographic information of the subjects used in the analysis to localize AD-vulnerable region.

	AD (N=188)	**Normal (N=169)**
Age mean±std	75.18±7.50	75.64±5.18
Gender M/F	99/89	83/86

2.1.2.2 WHICAP dataset

In a large well-characterized cohort from WHICAP study, we identified 7 cognitively normal APOE ε4 homozygous carriers from 383 cognitively normal subjects, who have stayed cognitively normal during the follow-up period. The demographic information can be found in Table 2-2. I carried out the analysis in a similar way as described in the previous section, where the testing variable is substituted to APOE ε4 homozygous carriers vs. others.

Table 2-2. Demographic information of the subjects from WHICAP study used to investigate the structural cortical characteristics of healthy APOE ε4 homozygous subjects.

	APOE ε4 homozygous carriers (N=7)	**Others (N=376)**
Age range	67.2 - 80.7	62.5 - 96.1
Age mean ± std	72.5 ± 5.0	73.7 ± 5.7
Gender M/F	4/3	156/220
APOE #ε4 2/1/0	7/-/-	-/121/255

2.1.2.3 Flat-map representation

In this study, I also proposed using the flat-map representation of the cortical surface (Bruce Fischl, Sereno, & Dale, 1999; David C Van Essen, Drury, Joshi, & Miller, 1998). Flat-map is the most compact way to represent the cortex, which provides unbiased visualization of the whole cortex in one single view, with no discontinuity or distortion (e.g. foreshortening) especially at the margin of the view. This representation facilitates observation and summarization of results, identification of clustered parcellation of interest (POI) and the comparison of the anatomical locations and spatial extents of the effects across different analyses. It also has computational merits because of the 2D representation. An illustration is shown in Figure 2-1.

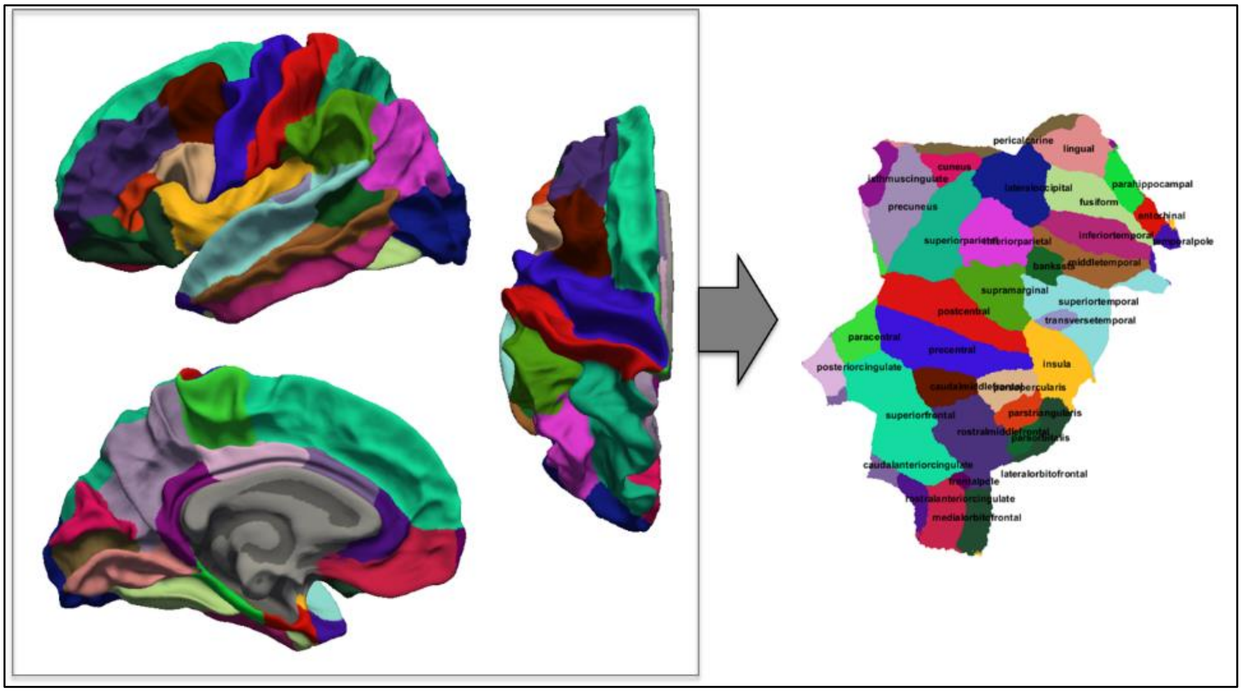


Figure 2-1. An illustration of flat-map representation of human cortex. (Left) Multiple 3D views are required for the full view of brain surface. (Right) Flat-map instead provides an unbiased single view of the whole cortex which facilitates identification of patterns and clusters.

2.1.3 Results

2.1.3.1 AD-vulnerable region

The results of the cortical thickness analysis of AD vs. cognitively normal subjects are summarized in Figure 2-2. In the flat-map representation, there presents a very focal AD-related atrophy in the anterior lateral entorhinal region. The affected region is rendered in other spaces as well. This points to the precise location within the full range of trans-entorhinal cortex (Augustinack et al., 2013) (indicated by the yellow arrows in the rightmost 3D surface space of Figure 2-2) that is especially vulnerable in AD pathology. And facilitated with the flat-map representation, there are other interesting observations, e.g. the involvement of temporal, frontal, and parietal lobes and the relative sparing of primary motor, sensory, visual cortices.

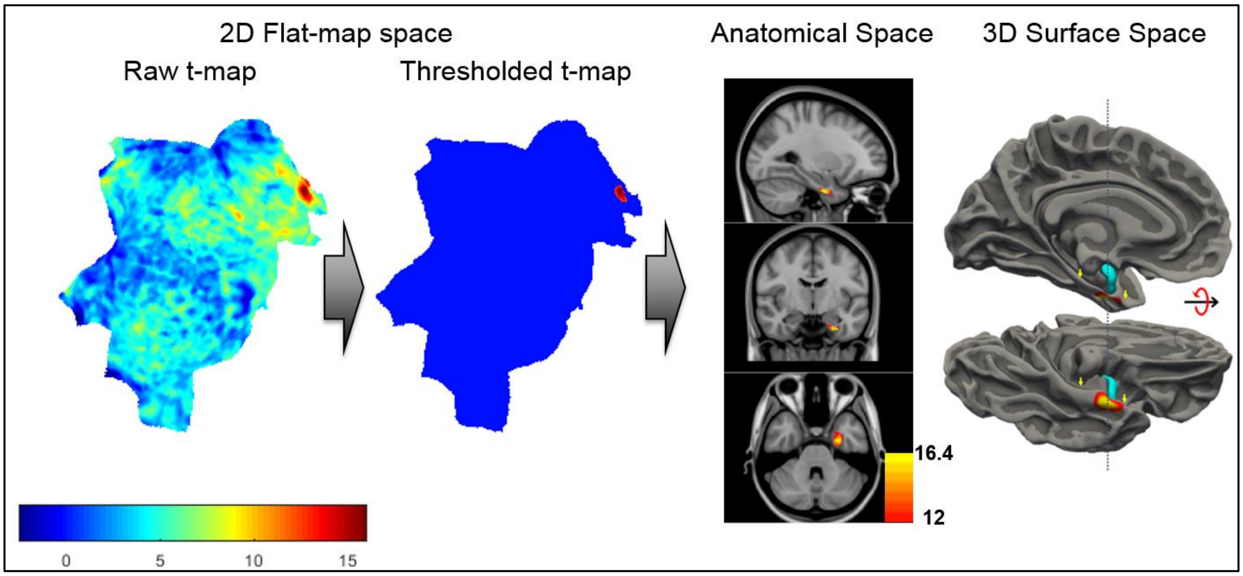


Figure 2-2. Cortical thickness analysis of AD vs. cognitively normal subjects. From left to right are the raw t-map and the thresholded t-map in the flat-map space, and the t-map in the anatomical space and the 3D surface space. In the 3D surface map, the cyan color indicates amygdala, the

yellow arrows point to the full range of transentorhinal cortex (Augustinack et al., 2013), the AD effect localizes to the anterior TEC at the level of amygdala.

2.1.3.2 Structural characteristics of healthy APOE ε4 homozygous subjects

The results of the cortical thickness analysis of healthy APOE ε4 homozygous carriers in the left hemisphere are summarized in Figure 2-3. There are two clusters with significant lower thickness that characterize the healthy APOE ε4 homozygous subjects: one is in superior frontal cortex, and the other is in anterior entorhinal cortex, as shown.

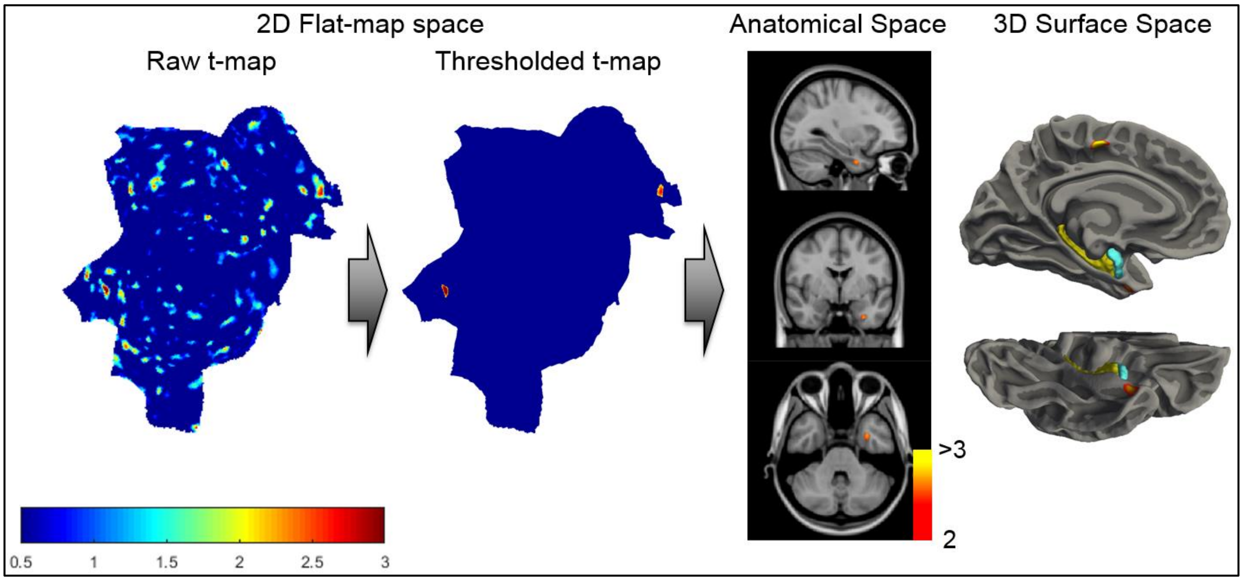


Figure 2-3. Cortical thickness analysis of healthy APOE ε4 homozygous carriers. From left to right are the raw t-map and the thresholded t-map at flat-map space, and the t-map in the anatomical space and 3D surface space. There are two significant clusters: one in superior frontal cortex, the other in anterior entorhinal cortex. In the 3D surface space, the cyan color indicates amygdala, the yellow color indicates hippocampus.

2.1.4 Discussion

2.1.4.1 AD-vulnerable region

The cortical thickness analyses pinpoint the transentorhinal cortex at the level of amygdala along the long axis that is prominent in AD pathology, see Figure 2-2. This observation helped provide evidence for further immunoblotting analysis in *ex vivo* human brain tissue samples, as shown in Figure 2-4.

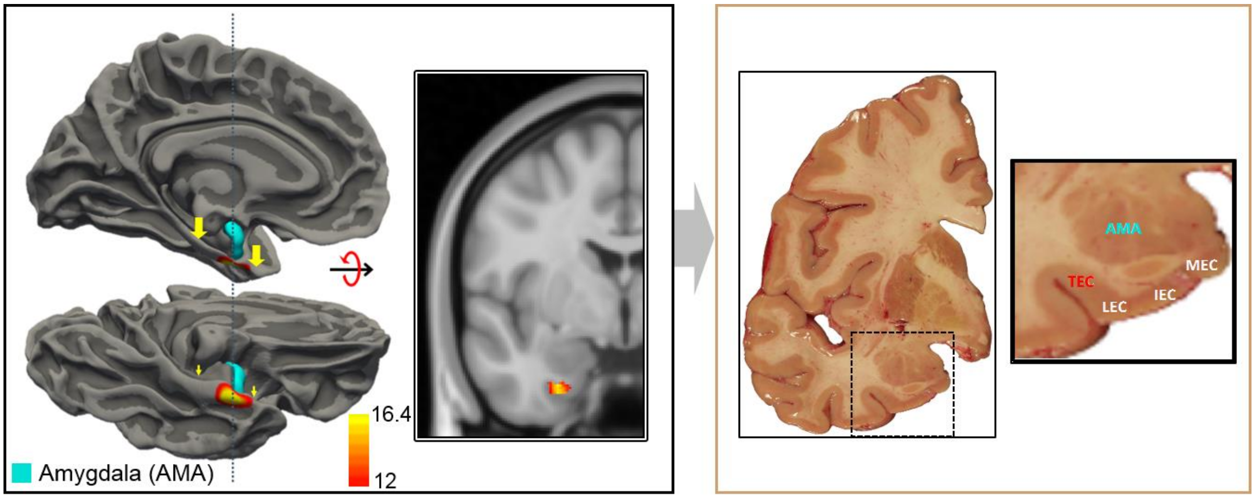


Figure 2-4. The neuroimaging finding helped guide molecular analysis of *ex vivo* brain tissue in AD-vulnerable region. (Left) Pixels that show the greatest and most reliable volumetric loss compared to controls are indicated in yellow/red. The TEC defect is shown on a coronal MRI slice. (Right) A representative human postmortem brain tissue slice, matching the precise anatomical coordinates of the neuroimaging finding, with a higher magnification illustrating the subregions of the entorhinal cortex isolated for protein profiling: the TEC, the lateral EC (LEC), intermediate EC (IEC) and medial EC (MEC).

2.1.4.2 Mechanistic implications from the healthy APOE ε4 homozygous subjects analysis

A previous study on healthy *middle age* subjects has shown significantly thinner cortex in superior frontal region of APOE ε3/4 group compared with ε3/3 group, with no significant effects of ε4 status on any temporal lobe measures (Fennema-Notestine et al., 2011). Whereas, a study in healthy *adolescents* reported thinner left entorhinal cortex in APOE ε4 carrier and suggested that the thinner entorhinal cortex in ε4 carriers might contribute to the risk of Alzheimer's disease (P. Shaw et al., 2007). In our analysis of *elderly* healthy APOE ε4 homozygous subjects, we observed significantly thinner regions in both superior frontal region and entorhinal cortex region. These results might imply complex genetic-age interaction.

For APOE gene, the regions with the highest expression levels are entorhinal cortex, parahippocampal cortex and temporal pole according to the data from Allen Brain Atlas[3] and the analytic method developed in (French & Paus, 2015). And a previous mouse study has reported ApoE4 protein exacerbates tau-mediated neurodegeneration (Shi et al., 2017). Taking into consideration the postmortem study of neurofibrillary tangles, where tauopathy starts from transentorhinal cortex (Braak & Braak, 1991), the observations might provide evidence explaining the risk of Alzheimer's disease in APOE ε4 carrier from a genetic-molecular co-localization point of view.

---

[3] http://human.brain-map.org

2.2 Normal aging

2.2.1 Introduction

Extending prior studies in normal aging, we performed a more thorough functional and structural investigation of normal aging using data from an aging study our lab recently completed, in which both CBV and volume were mapped in over 100 carefully-screened cognitively normal individuals spanning 20-72 years of age (Figure 2-6). This cohort will be denoted as Columbia cohort hereinafter in this chapter. We analyzed the age-related changes in both CBV and volume.

We further relied on a mouse study to establish an MRI-based profile reflective of dendritic spine loss, and provide mechanistic evidence for the findings in the human normal aging study.

We further set out to broaden the age-span to older subjects, which is important for understanding how the brain is affected throughout the full age span. We relied on longitudinal clinical follow-up and CSF biomarker data to ensure the study population is free of AD.

Collectively, the studies show that the dentate gyrus within the hippocampus and the inferior frontal gyrus across the cortex are the two brain regions differentially affected by normal aging structurally and functionally and fulfil the MRI-based profile of dendritic spine loss, suggesting dendritic spine loss as the underlying pathophysiology. The entorhinal cortex, a region most vulnerable to AD, turned out to be the region most resistant to normal aging.

2.2.1.1 Concordant changes in functional and structural MRI reflect dendritic spine loss

Both functional and structural neuroimaging should be sensitive to dendritic spine loss in normal aging, since dendritic spines are known to have high metabolic needs (Z. Li, Okamoto, Hayashi, & Sheng, 2004). This is validated by various functional imaging studies in both *cerebral cortex* (de Leon et al., 2001; Gazzaley et al., 2005; Martin, Friston, Colebatch, & Frackowiak, 1991) and *hippocampus* (Brickman et al., 2014; Pavlopoulos et al., 2013; Scott A Small, Chawla, Buonocore, Rapp, & Barnes, 2004; Yassa, Mattfeld, Stark, & Stark, 2011). Structural imaging is also expected to be sensitive to spine loss because of the space-occupying spine-related cell elements. Studies have shown that variables derived from structural MRI are sensitive to dendritic spine loss (Kassem et al., 2013; Keifer Jr et al., 2015).

While both functional and structural imaging could be sensitive to dendritic spine loss, they are not necessarily specific. This is particularly the case for structural measures, as there are numerous elements that can contribute to volumetric changes independent of spine loss. And neuronal death is a neuropathological process that can drive decreases in both functional and structural measures. We hypothesized that, in the absence of neuronal death, a concordant change in functional and structural measures might represent a neuroimaging profile reflective of dendritic spine loss. This hypothesis was validated in an unilateral whisker cutting experiment in young mice led by Dr. Jia Guo, where the intervention induces dendritic spine loss in select barrel cortex layers (Y. Zuo, Yang, Kwon, & Gan, 2005).

In the mouse study, CBV in each layer was measured in two ways. The first is %CBV, which is an estimate of CBV density within a unit volume of brain, similar to density estimates used in prior functional imaging studies (de Leon et al., 2001; Khan et al., 2014). However, as established for other functional readouts of the brain, for example neurons or synapses, density measures introduce potential biases (Kempermann, Kuhn, & Gage, 1997; Morrison & Hof, 1997; Napper, 2018; West, 1999). When possible, “total” measurements are proven superior, typically generated by multiplying density and regional volume (Kempermann et al., 1997; West, 1999). We therefore used this approach to also generate total CBV (tCBV) for each layer. Only layers containing synapses have a loss in both CBV and volume in the contralateral barrel cortex after unilateral whisker cutting, and have a correlation between both measures. And as anticipated, tCBV was found superior to %CBV in mapping these changes.

Together, we concluded that when it is known that a condition is not associated with neuronal death, as is the case with normal aging (Morrison & Hof, 1997), an MRI-based profile reflective of dendritic spine loss should fulfill two criteria: a reduction in regional tCBV and regional volume, and a correlation between both MRI-based variables.

We relied on these insights from the mouse findings to map correlates of dendritic spine loss in normal aging in human. We found dentate gyrus and the inferior frontal gyrus were both differentially affected by normal aging structurally and functionally and showed concordance, suggesting dendritic spine loss as the underlying neurophysiological mechanism of the age-related changes in these two regions.

2.2.1.2 Necessity and method to exclude the confound of AD

For normal aging studies, one challenge has been excluding AD, especially preclinical AD which occurs years before its clinical onset and is usually undetectable by conventional clinical measures (Sperling et al., 2011). The AD confound is particularly acute when attempting to cover the tail end of the age-span. Preclinical AD's prevalence is estimated to begin its dramatic age-related rise during the 8th - 9th decades of life (Brookmeyer, Abdalla, Kawas, & Corrada, 2018), with the estimated prevalence of preclinical AD reaching as high as 50%.

Only a few subjects of the Columbia cohort were in the high prevalence age range, and the fact that the entorhinal cortex was found to be unaffected by aging, supports the interpretation that the Columbia cohort was not dominated by subjects with occult AD dementia.

To investigate an older cohort in whom the AD confound can be excluded, we turned to the ADNI. ADNI includes a large population of cognitively normal participants who have been followed longitudinally and who have, among extensive clinical and neuropsychological testing, CSF biomarkers of AD. While ADNI does not acquire high-resolution functional imaging data, it does acquire high resolution structural MRI. We derived volumetric measurements in participants who met strict clinical and CSF biomarker criteria against AD. Two CSF cutoffs were established by ADNI (L. M. Shaw, Figurski, Waligorska, & Trojanowski, 2016), one that relies on CSF Aβ alone and the other that relies on a combination of CSF tau and Aβ. We opted for the later, as recent studies have shown that AD-associated elevations in CSF tau, and tau pathology itself, are strong determinants of regional dysfunction and atrophy (Tarawneh et al., 2015; Thaker et al., 2017). In addition, we also performed analyses using tau/Aβ as a continuous covariate instead of setting a binary grouping using the cut-off value.

2.2.2 Methods

2.2.2.1 Columbia cohort

For Columbia cohort, subjects are healthy adults between age 20 and 72, recruited from the Columbia University Medical Center/New York Presbyterian Hospital campus. MRI acquisition was part of a Columbia University Medical Center Institutional Review Board approved study, with explicit written consent obtained from each participant. The data selection process is illustrated in the left panel of Figure 2-5, with the demographic information of the CBV-fMRI dataset summarized in Figure 2-6, the demographic information of the structural MRI dataset summarized in Figure 2-7.

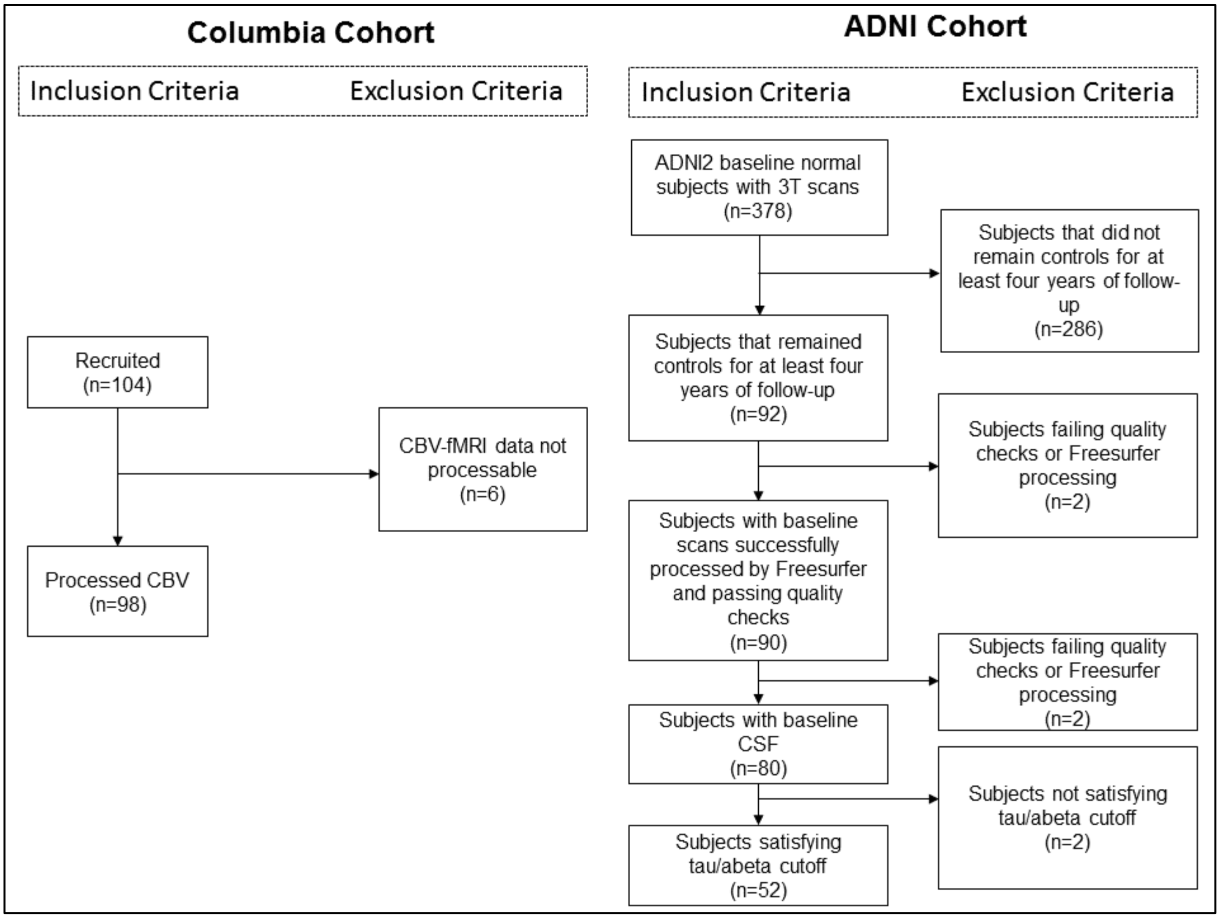
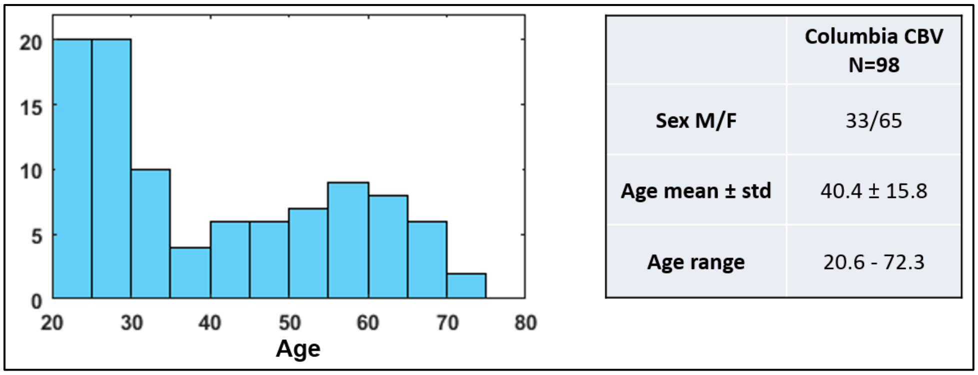


Figure 2-5. Dataset selection process for the Columbia and ADNI cohorts in the normal aging study.



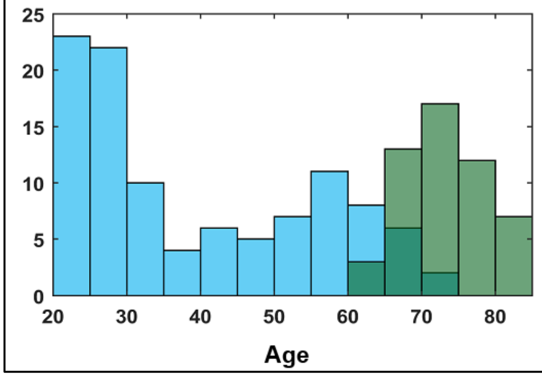
	Columbia CBV N=98
Sex M/F	33/65
Age mean ± std	40.4 ± 15.8
Age range	20.6 - 72.3

Figure 2-6. Demographic information of the subjects in the CBV-fMRI normal aging study.

2.2.2.2 Cross-sectional cohort of normal aging elderlies from ADNI

As described in the introduction of the chapter, we adopted two strategies to deal with the AD confound as indicated best by the composite of CSF tau/Aβ ratio. From ADNI, we identified 80 subjects who were cognitively normal at baseline and stayed normal during at least 4 years of clinical follow-up and who have baseline CSF biomarker. In the first analysis, we only included the 52 subjects who were CSF biomarker negative for AD, using the combined tau/Aβ cutoff (tau/Aβ < 0.39) (L. M. Shaw et al., 2016; L. M. Shaw et al., 2009). The demographic information of the 52 subjects is summarized in

Figure **2-7**. In another analysis, we included all 80 subjects and included the tau/Aβ ratio as a covariate. The data selection process is illustrated in the right panel of Figure 2-5.



	Columbia Volume N=104	ADNI N=52
Sex M/F	36/68	30/22
Age mean ± std	39.9 ± 15.9	72.9 ± 5.9
Age range	20.6 - 72.3	62.4 - 84.5

Figure 2-7. Demographic information of the subjects in the structural normal aging study covering 20-85 years old.

2.2.2.3 Longitudinal cohort of normal aging elderlies from ADNI

From the 80 subjects in the cross-sectional analysis, we further identified 62 subjects who have five longitudinal MRI scans at baseline, month 6, 12, 24, and 48 for longitudinal analysis. In the similar manner as the cross-sectional analysis, we identified 40 subjects who were CSF biomarker negative for AD using the tau/Aβ cutoff. And additionally, we included all 62 subjects and included the tau/Aβ ratio as a covariate in the analysis.

2.2.2.4 MRI acquisition

For the CBV-fMRI, we used a steady-state contrast enhanced CBV technique as previously described (Brickman et al., 2014; Khan et al., 2014). MRI scans were acquired with a Philips Achieva 3.0 T MRI scanner using an 8-channel SENSE head coil. In each scan session, a T1-weighted structural scan (TR = 6.7 ms, TE = 3.1 ms, FOV = 240 × 240 × 192 mm^3, voxel size = 0.9 × 0.9 × 0.9 mm^3) was first acquired using a Turbo Field Echo (TFE) gradient echo (GRE) sequence; a pair of un-scaled T1-weighted images (TR = 7 ms, TE = 3 ms, FOV = 240 × 240 × 162 mm^3, voxel size = 0.68 × 0.68 × 3 mm^3) were acquired afterwards with a bolus injection of gadolinium contrast agent in between.

2.2.2.5 MRI processing

The T1-weighted structural images were processed using FreeSurfer, generating cortical parcellation (Desikan et al., 2006; B. Fischl et al., 2004) and hippocampal subregions segmentation (Iglesias et al., 2015) in the individual space. The list of cortical regions can be found in the parcellation protocol documented in (Desikan et al., 2006). The primary hippocampal subregions labeled include presubiculum (PRESUB), subiculum (SUB), CA1, CA3, CA4 (hilus), granule cell molecular layer of DG (DG), molecular layer of subiculum and CA fields (MLSUBCA).

CBV-fMRI processing followed the steps applied in previous studies (Brickman et al., 2014; Khan et al., 2014) and included registration of the pre-contrast and the post-contrast T1-

weighted scans, subtraction of the co-registered post-contrast and pre-contrast scans, and CBV value normalization with the mean signal in the superior sagittal sinus. The raw CBV values are %CBV measures in a unit voxel.

Individual structural images were registered into template space with a symmetric diffeomorphic algorithm using ANTS (Avants et al., 2011). Individual CBV images were linearly registered into the individual structural image space. The CBV images were registered to the template space with the transformation field composed of the linear transformation matrix and the diffeomorphic transformation field.

In ROI analyses, total CBV (tCBV) measures the total amount of cerebral blood volume in an anatomically defined ROI and was calculated as the product of the regional %CBV density and the ROI volume. In voxel/vertex-based analyses, tCBV measures the voxel/vertex-wise local total amount of cerebral blood volume and is the voxel/vertex-wise product of CBV and volumetric measures.

2.2.2.6 Image Analysis

2.2.2.6.1 Total CBV voxel-based analysis in the hippocampus

For each voxel, the tCBV value is the product of %CBV and the Jacobian determinant, which measures the relative volume change warping from unit volume element into the template space. Voxel-wise linear regression with age as regressor, gender and ICV as covariates was performed using SPM8[4]. The results were thresholded at $p < 0.05$ and spatial cluster size larger than 10.

---

[4] http://www.fil.ion.ucl.ac.uk/spm/software/spm8

2.2.2.6.2 Total CBV vertex-based analysis in the cortex

Vertex-based analysis was used for cortical tCBV analysis. Individual CBV-fMRI images were registered into individual structural image space using linear registration. The cortical surface in individual structural image space was reconstructed using FreeSurfer (Dale, Fischl, & Sereno, 1999). The co-registered CBV-fMRI image was projected onto the reconstructed cortical surface along the surface normal with projection fractions from 0 to 1 in the step of 0.1. The projected maps were then averaged into a single map generating robust measures of surface-based cortical CBV. The individual surfaces along with the projected CBV maps were registered into the FreeSurfer *fsaverage* space for inter-subject analysis. Cortical volume maps in the same space were generated using FreeSurfer, where the volume measure at each vertex is the volume of the local truncated tetrahedron. The tCBV at each vertex is the product between the %CBV and volume. Vertex-wise linear regression with age as regressor, gender and ICV as covariates, was performed using custom scripts. The results were thresholded at $p < 0.01$ and spatial cluster size larger than 500.

2.2.2.6.3 ROI analysis

The cortical parcellation and hippocampal subregion segmentation label maps were registered into the CBV-fMRI image space with the transformation matrix from registering individual CBV image into the individual structural image space. Regional %CBV is the average %CBV within each ROI. Regional tCBV is the product of the regional %CBV and the ROI volume. Linear regression with age as regressor, gender and ICV as covariates was performed for each ROI. It has to be noted that the frontal pole labeled in the segmentation protocol was not fully covered in a small portion of the CBV-fMRI scans because of the field of view (FOV) thus were not included in the results. In the volume ROI analyses, for each ROI, I used linear regression with ROI volume as dependent variable, age as regressor, gender and ICV as covariates.

2.2.2.6.4 Longitudinal ROI analysis

In the longitudinal volume ROI analysis, I performed repeated measure ANOVA with baseline age, gender, ICV, and baseline tau/Aβ ratio as covariates; and in the other analysis with subjects below tau/Aβ cut-off, I performed repeated measure ANOVA with baseline age, gender, ICV as covariates. The p-value of the effect of time on regional volumes with lower bound adjustment were reported.

2.2.3 Results

In the exploratory analysis, the most reliable age-related changes were observed in the left brain, similar to what has been previously described (Chetelat et al., 2013; de Leon et al., 2001; Martin et al., 1991). We, however, cannot exclude technical reasons for this lateralizing effect. All analyses were conducted on measurements derived from the left brain.

2.2.3.1 Brain regions vulnerable and resistant to normal aging from 20-72 years of age

2.2.3.1.1 CBV-fMRI

Results revealed that across the cortex the most reliable age-related tCBV decrease was observed in the vicinity of the inferior frontal gyrus (Figure 2-8A). For the complimentary ROI analysis, the only two regions that showed significant age-related CBV decreases at significance level $\alpha = 1e\text{-}6$, after applying a Šidák correction, were two constituent subregions of the inferior frontal gyrus (pars orbitalis: $t = -6.28$, $p = 1.04e\text{-}8$; pars triangularis: $t = -6.08$, $p = 2.61e\text{-}8$, two-tailed, $N = 98$) (Figure 2-8B). Unexpectedly, the entorhinal cortex was the cortical region that was least affected by aging (t = 0.411, p = 0.682, two-tailed, N = 98) (Figure 2-8B). Within the hippocampus, the most reliable age-related decline was observed in the dentate gyrus (Figure 2-8C). For the ROI analysis, the dentate gyrus was found to be the dominant site of age-related CBV decline (t = -6.19, p = 1.53e-8, two-tailed, N = 98) (Figure 2-8D).

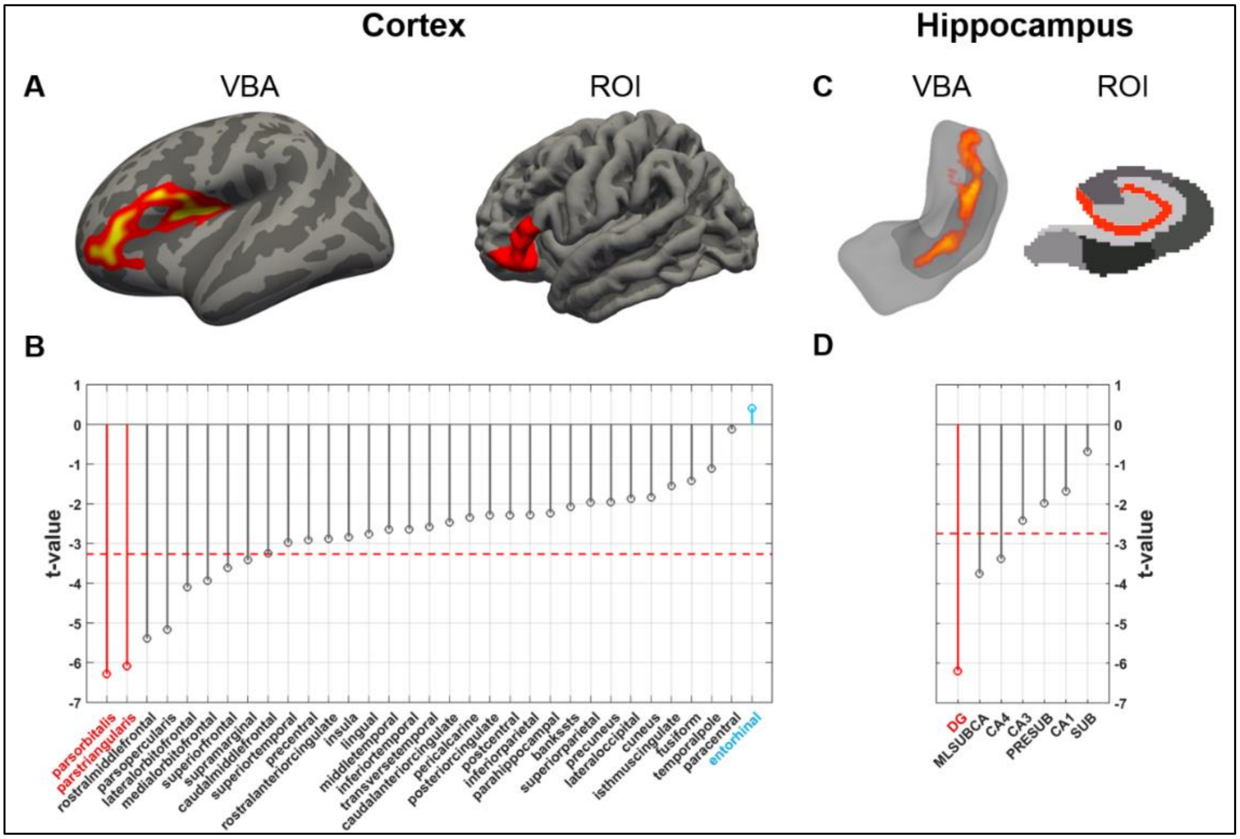


Figure 2-8. Mapping normal aging with CBV-fMRI from 20-72 years of age. (A) A vertex-based analysis of the cortex (VBA; left image) and a region-of-interest analysis across cortical regions (ROI; right image) identified the greatest age-related decline of cerebral blood volume (CBV) in the inferior frontal gyrus. (B) The t-value distribution of age-related CBV decline across cortical regions shows that two regions of the inferior frontal gyrus (indicated in red, the pars orbitalis and the pars triangularis) are most reliably vulnerable to aging. The entorhinal cortex (indicated in blue) was found most resistance to aging. The dashed red line indicates the t-value

threshold at α = 0.05 adjusted for Šidák multiple comparison. (C) A voxel-based analysis of the hippocampus (VBA; left image) and a region-of-interest analysis across hippocampal regions (ROI; right image) identified the greatest age-related CBV decline in the dentate gyrus. (D) The t-value distribution of age-related CBV decline across hippocampal regions, shows that the dentate gyrus (indicated in red) is most reliably vulnerable to aging. The dashed red line indicates the t-value threshold at α = 0.05 adjusted for Šidák multiple comparison.

The corresponding analysis results using percentage CBV are shown in **Figure 2-9**, revealing similar localization results.

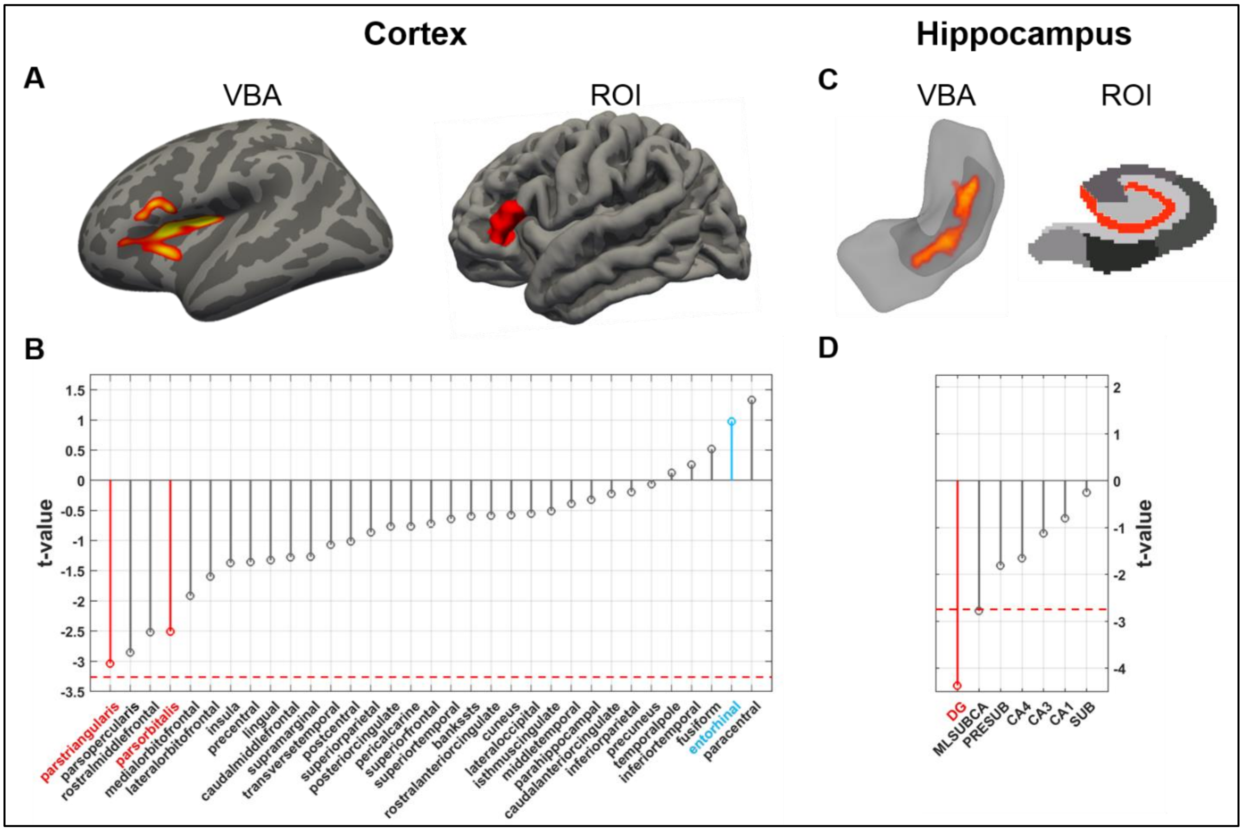


Figure 2-9. Mapping aging with CBV-fMRI from 20-72 years of age using percentage CBV.

2.2.3.1.2 Volumetric MRI

For volumetry results, across the cortex, subregions of the inferior frontal gyrus were found significantly associated with aging (pars orbitalis: t = -7.50, p = 2.62e-11; pars triangularis: t = -6.88, p = 5.38e-10, two-tailed, N = 104), although, unlike for CBV, their volumes were not the ones most reliably affected by aging (Figure 2-10). As with CBV, the entorhinal cortex was the cortical region least affected by aging (t = -0.445, p = 0.657, two-tailed, N = 104). Within the hippocampus, the dentate gyrus was again the region differentially affected by aging (t = -5.04, p = 2.08e-6, two-tailed, N = 104).

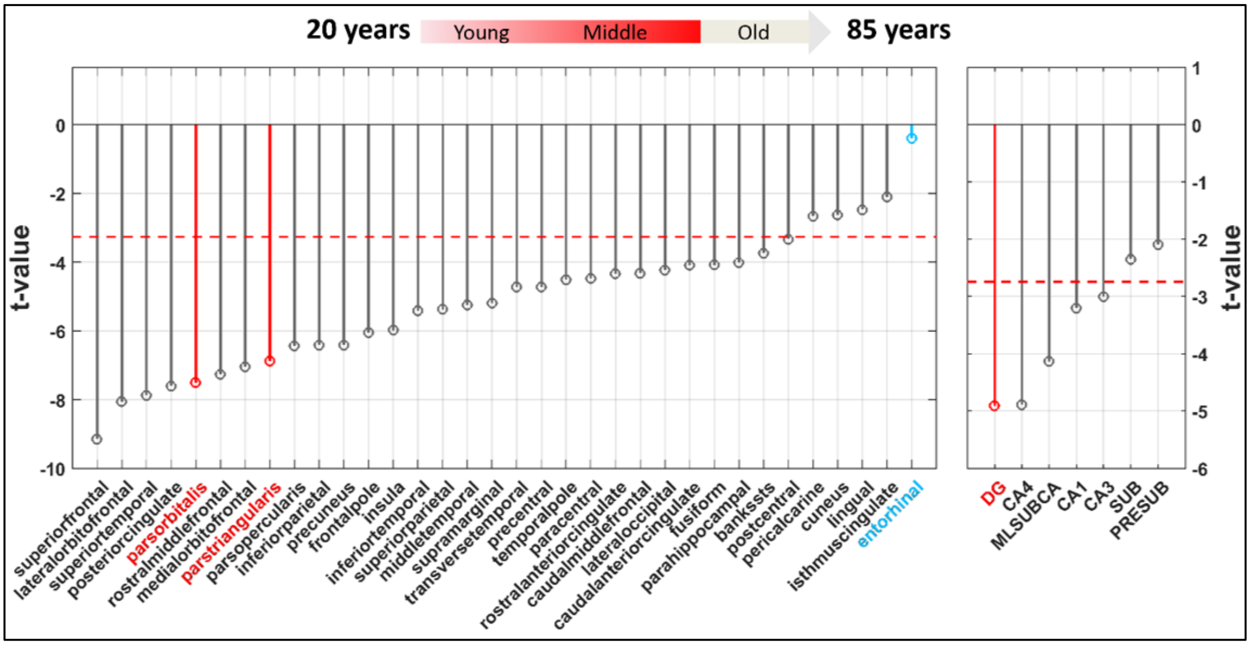


Figure 2-10. Mapping normal aging with volumetric MRI from 20-72 years of age. The t-value distribution of age-related decline in volume, measured by structural MRI, across cortical regions. While not most reliably affected, volumes of regions of the inferior frontal gyrus (indicated in red) decline significantly with age. The volume of the entorhinal cortex (indicated in blue) was found most resistance to aging. The dashed red line indicates the t-value threshold at α = 0.05 adjusted for Šidák multiple comparison.

2.2.3.1.3 CBV-Volume relationships

Guided by the mouse study, we found a significant correlation between tCBV and volume in the dentate gyrus (r = 0.524, p = 3.14e-8, two-tailed, N = 98) and the inferior frontal gyrus (r = 0.563, p = 2.00e-9, two-tailed, N = 98) (Figure 2-11), and so both regions fulfill the criteria for MRI-based profile reflective of dendritic spine loss. In an exploratory analysis, we found that some regions fulfilled one criterion but not both, e.g. the paracentral lobule showed significant age-related volume reduction (t = -4.48, p = 2.13e-5, two-tailed, N = 98) but not tCBV reduction (t = -0.123, p = 0.902, two-tailed, N = 98) or a reliable correlation between tCBV and volume (r = 0.054, p = 0.601, two-tailed, N = 98) (Figure 2-11). The correlation values of the whole set of regions can be found in **Figure 2-12**.

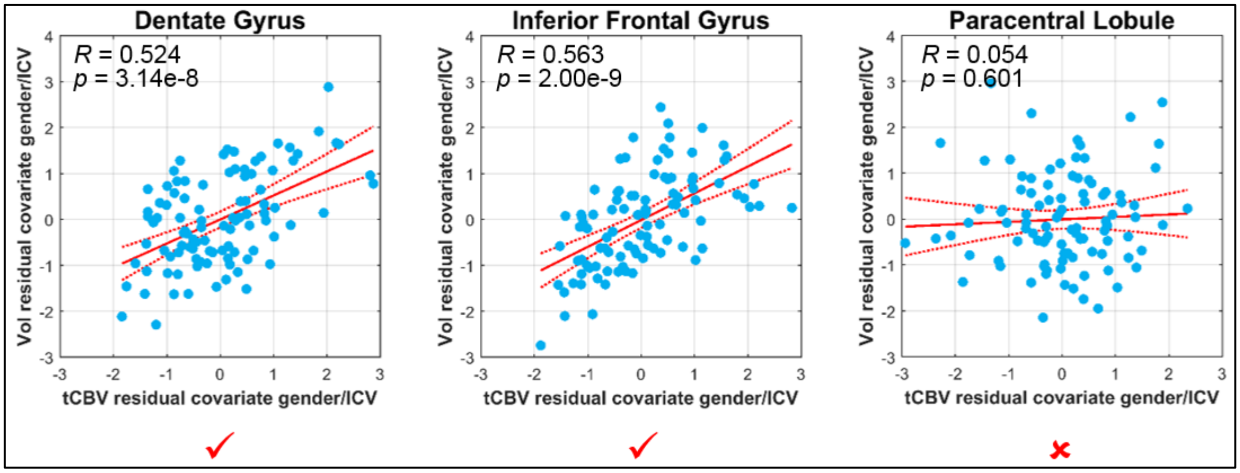


Figure 2-11. The relationship between volumetric MRI and CBV-fMRI. A significant concordant relationship between CBV and volume is observed for the dentate gyrus and the inferior frontal gyrus, consistent with the MRI profile of dendritic spine loss, but not for the paracentral lobule as an example.

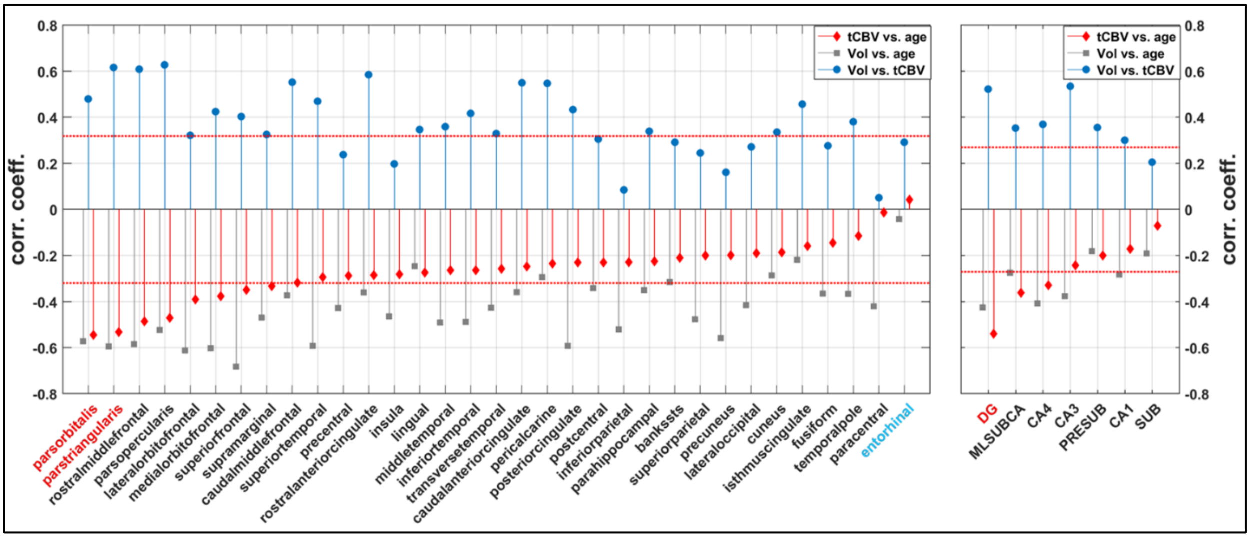


Figure 2-12. A summary graph of the partial correlation coefficients of tCBV versus age, volume versus age, and tCBV versus volume. Gender and ICV were included as covariates. The dashed red line indicates the correlation coefficient threshold at *α* = 0.05 adjusted for Šidák multiple comparison.

2.2.3.2 Brain regions vulnerable and resistant to normal aging from 62-85 years of age

We generated cortical and hippocampal volumes as with the Columbia cohort and then applied the same linear regression model to each volume (Figure 2-13). Across the cortex, regions of the inferior frontal gyrus were found unaffected by aging in this age-span, but here again the entorhinal cortex was one of the regions least affected by aging. Within the hippocampus, the dentate gyrus was again the region most reliably affected by aging (t = -2.37, p = 0.0217, two-tailed, N = 52), however, likely because of the constrained age-range, the effect in the dentate gyrus did not meet the Šidák-corrected threshold for significance at α = 0.05. The regional pattern was similar in the analysis with the inclusion of tau/Aβ ratio as a covariate (Figure 2-14).

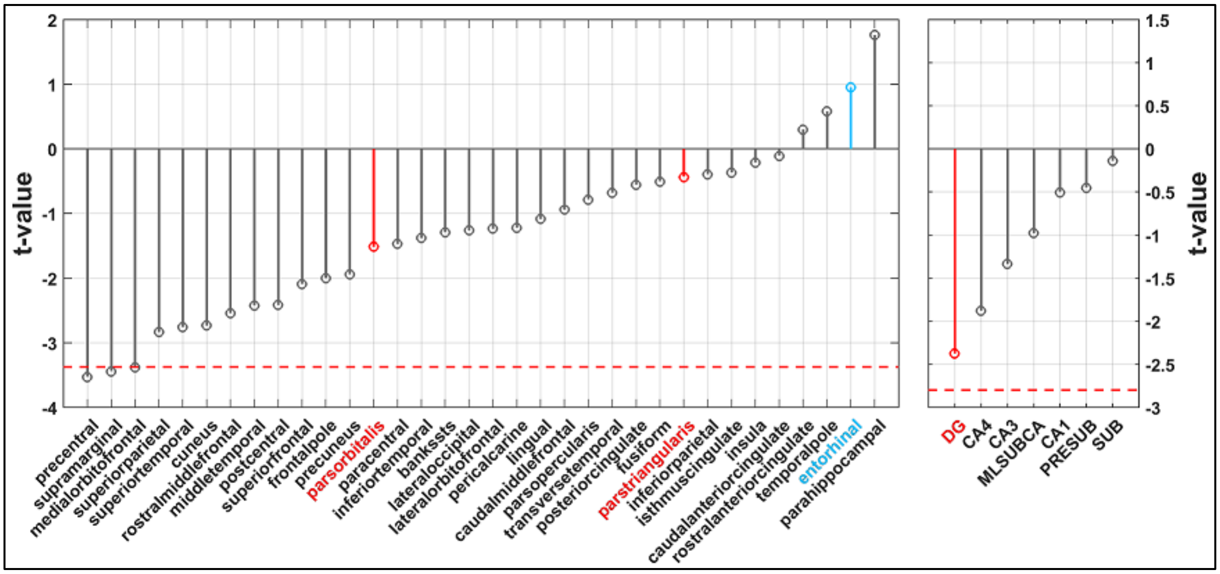


Figure 2-13. Mapping normal aging with volumetric MRI in AD-free subjects from 62 - 85 years of age. The t-value distribution of age-related decline in volume across cortical and hippocampal region, shows that the dentate gyrus (indicated in red, right graph) is most vulnerable to aging and the inferior frontal gyrus (indicated in red, left graph) is not reliably associated with aging. The entorhinal cortex (indicated in blue, left graph) is least affected by aging. The dashed red line indicates the t-value threshold at α = 0.05 adjusted for Šidák multiple comparison.

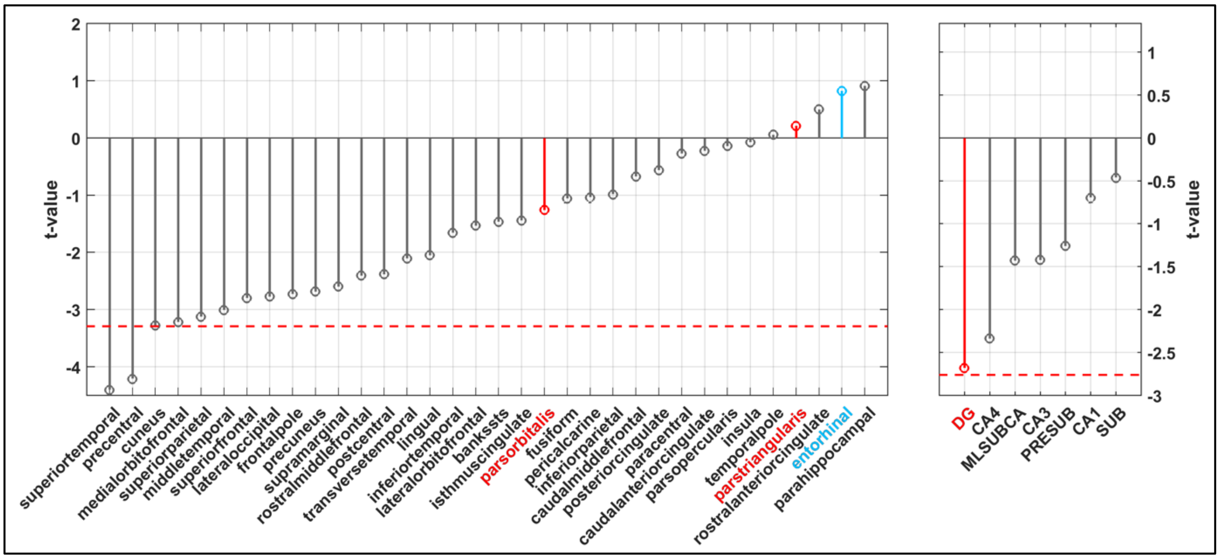


Figure 2-14. Mapping normal aging with volumetric MRI from 62-85 years of age with tau/Aβ as covariate. The t-value distribution of age-related volume decline across cortical and hippocampal region with tau/Aβ as covariate, from 62-85 years of age, shows that the dentate gyrus (indicated in red, right graph) is most vulnerable to aging (although not crossing threshold of multiple comparisons) and the inferior frontal gyrus (indicated in red, left graph) is not reliably associated with aging. The entorhinal cortex (indicated in blue, left graph) is among regions least affected by aging. The dashed red line indicates the t-value threshold at α = 0.05 adjusted for Šidák multiple comparison.

2.2.3.3 Longitudinal analysis of brain regions vulnerable and resistant to normal aging

No region, including entorhinal cortex and dentate gyrus, showed significant longitudinal atrophy when controlling for the covariates (Figure 2-15, Figure 2-16).

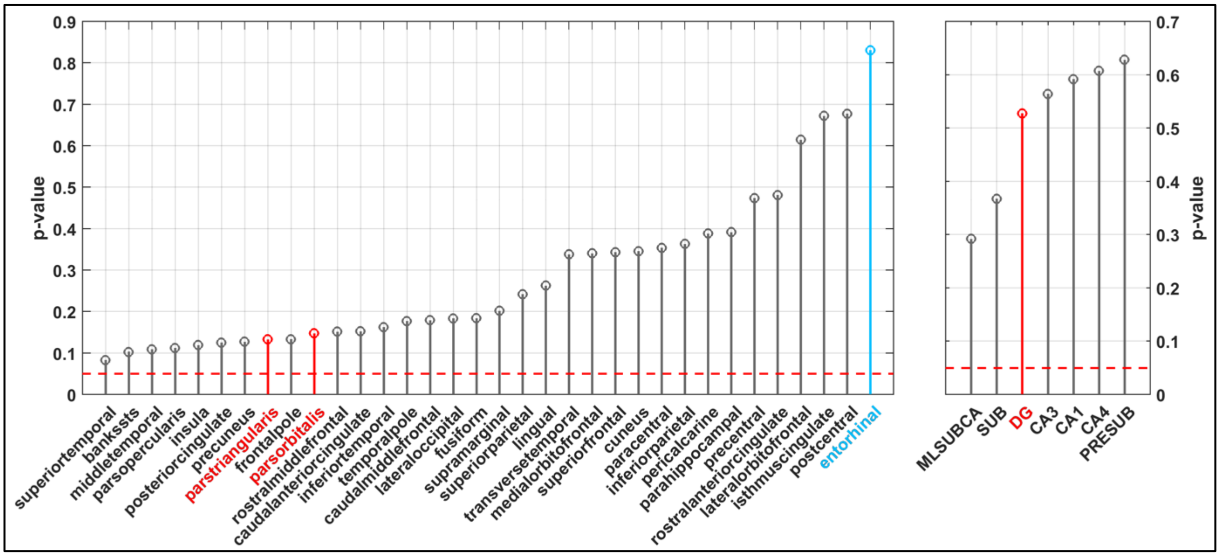


Figure 2-15. Mapping longitudinal age-related change with volumetric MRI from 62-85 years of age with tau/Aβ as covariate. The p-value distribution of longitudinal age-related

volume decline across cortical and hippocampal regions with tau/Aβ as covariate, from 62-85 years of age, illustrates that no region is showing reliable longitudinal age-related atrophy. The dashed red line indicates α = 0.05.

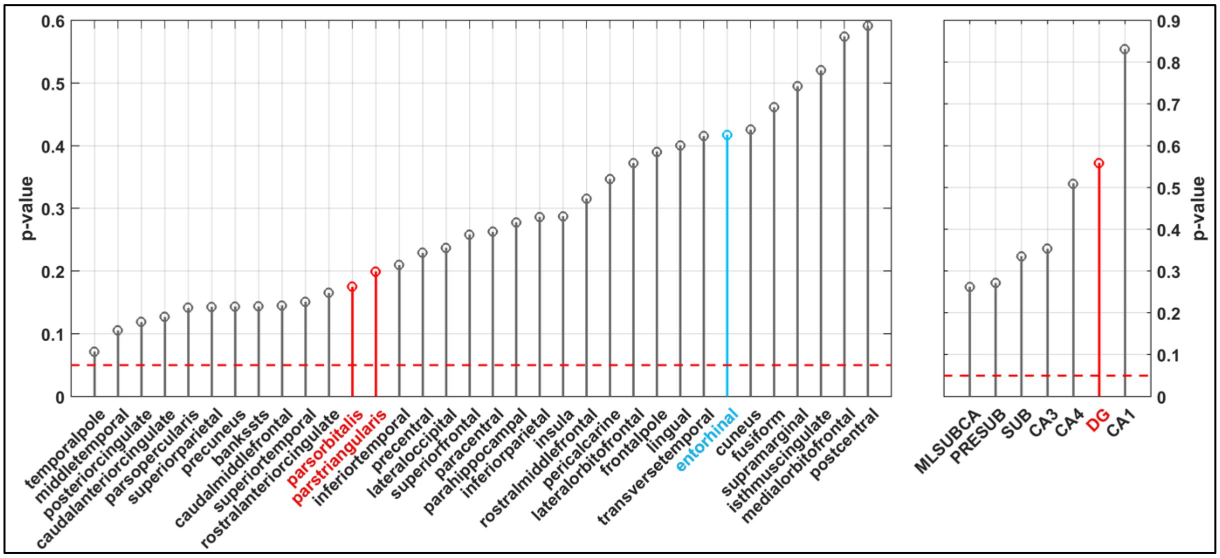


Figure 2-16. Mapping longitudinal age-related change with volumetric MRI from 62-85 years of age using tau/Aβ cutoff. The p-value distribution of longitudinal age-related volume decline across cortical and hippocampal regions with baseline tau/Aβ under cut-off threshold, from 62-85 years of age, illustrates that no region is showing reliable longitudinal age-related atrophy. The dashed red line indicates α = 0.05.

2.2.3.4 Brain regions vulnerable and resistant to normal aging across age span

To map aging independent of AD across the full adult lifespan — from young adulthood, through midlife, to old age, we merged the Columbia cohort with the ADNI cohort, as shown in

Figure **2-7**. By mapping the age-related volumetric changes (Figure 2-17), the most remarkable result is that among all brain regions investigated, the entorhinal cortex is the region most resistance to normal aging (t = 1.21, p = 0.227, two-tailed, N = 156). Not surprisingly, based on the previous analyses, the dentate gyrus was the hippocampal region most reliably affected by aging (t = -6.61, p = 6.24e-10, two-tailed, N = 156). The inferior frontal gyrus, while not the region most affected, was still found significantly affected by aging (pars orbitalis: t = -9.63, p = 1.99e-17; pars triangularis: t = -7.67, p = 1.96e-12, two-tailed, N = 156).

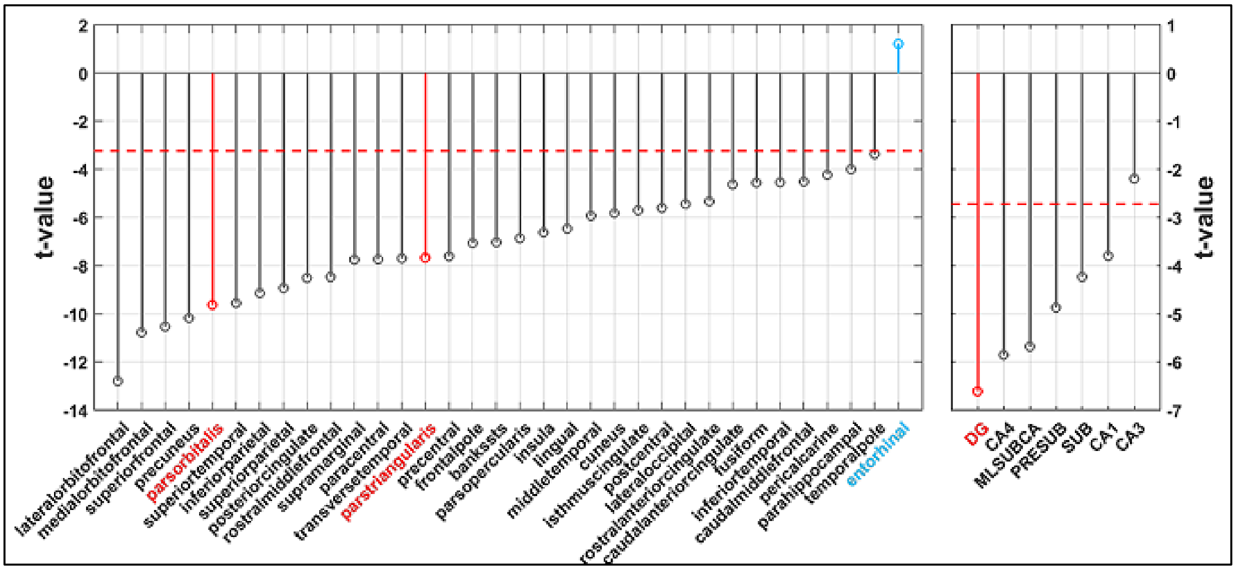


Figure 2-17. Mapping normal aging with volumetric MRI across the adult age-span. The t-value distribution of age-related decline in volume across cortical and hippocampal region in AD-free subjects across the full age-span, shows that the dentate gyrus (indicated in red, right graph) is most vulnerable to aging and the inferior frontal gyrus (indicated in red, left graph) is reliably associated with aging. The entorhinal cortex (indicated in blue, left graph) is the region least affected by aging. The dashed red line indicates the t-value threshold at α = 0.05 adjusted for Šidák multiple comparison.

We relied on the combined dataset to explore the trajectories of the dentate gyrus, the inferior frontal gyrus, and the entorhinal cortex across the full age-span. Interestingly, the vulnerable regions appear to have different aging trajectories. While the age-related effect of the dentate gyrus worsens linearly across the full age-span (p_{age} = 1.25e-8), the inferior frontal gyrus was better modeled by a quadratic curve (p_{age} = 3.79e-4, $p_{age\wedge 2}$ = 1.81e-2) (Figure 2-18 A) than a linear model, where the effect of aging on this brain region appears to taper off during normal aging. The Akaike information criterion (AIC) of the quadratic regression model is lower than the linear regression model. The entorhinal cortex was best modelled as a flat line across the full age span (p_{age} = 0.261) (Figure 2-18 A).

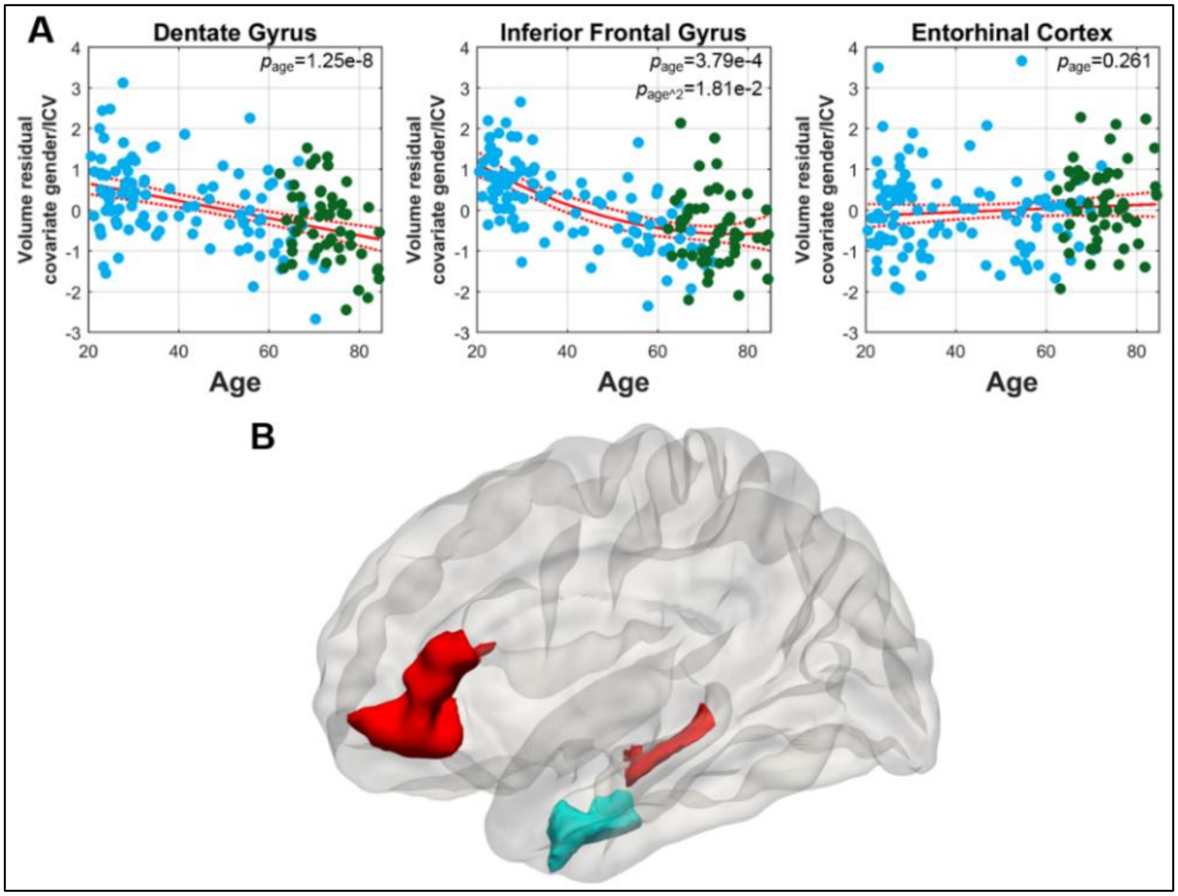


Figure 2-18. Trajectories of brain regions vulnerable and resistant to normal aging across the age-span. (A) The aging trajectory of dentate gyrus volume (left image) shows a linear decline across the age-span. The trajectory of inferior frontal gyrus volume (middle image) shows a

curvilinear decline. The trajectory of entorhinal cortex volume (right image) shows that it is unaffected by aging across the age-span. (B) A graphic summary of the two regions differentially vulnerable to normal aging, the dentate gyrus and the inferior frontal gyrus (red), and the region most resistant to normal aging, the entorhinal cortex (blue).

2.2.3.5 Cross-sectional volumetric analyses in independent life-span normal aging studies

I also performed cross-sectional volumetric analyses in three independent normal aging studies covering full adult age-span: Cam-CAN (Taylor et al., 2017), IXI, DLBS (Rodrigue et al., 2012). The results show similar patterns as those revealed in our cohort: dentate gyrus shows the most reliable age-related decrease within hippocampus, whereas the entorhinal show little age-related difference, IFG shows age-related decrease but superior frontal region shows the most reliable age-related decrease. The detailed results are included in the appendix A.1.

2.2.4 Discussion

In hippocampal formation, dentate gyrus was observed to be differentially vulnerable to normal aging. This agrees with previous imaging studies in non-human mammals (Moreno et al., 2007; Scott A Small et al., 2004), who develop age-related hippocampal dysfunction without the confound of AD, and also agrees with a recent post-mortem analysis of the disease-free human hippocampus (Adler et al., 2018). Across the cortex, the inferior frontal gyrus is the region most reliably affected by normal aging. This too agrees with previous studies (de Leon et al., 2001; Herholz et al., 2002; Kalkstein, Checksfield, Bollinger, & Gazzaley, 2011; Martin et al., 1991).

Since we used a high-resolution CBV-fMRI technique, we were able to map, for the first time, age-related dysfunction in the dentate gyrus and inferior frontal gyrus in the same cohort. More importantly, since both the dentate gyrus and the inferior frontal gyrus fulfilled MRI-based criteria for spine loss, our findings provide primary evidence that age-related dendritic spine loss is the likely driver of dysfunction.

Both cohorts and the combined analysis provide definitive evidence that within the hippocampus the dentate gyrus is the region most dominantly affected by normal aging (Figure 2-18). Findings from the ADNI cohort and the combined analysis suggest that age-related inferior frontal gyrus dysfunction tapers off in older age (Figure 2-18). However, as this observation was based only on volumetric information, we are less certain about this result. In both cohorts and in the combined analysis, the entorhinal cortex turned out to be the brain region differentially resistant to normal aging (Figure 2-18). We consider this a remarkable and unexpected finding, as this is the same region that is differentially vulnerable to AD (Braak & Braak, 1991; Gómez-Isla et al., 1996; Khan et al., 2014; Miller et al., 2015). Our study is not the first to suggest that the entorhinal cortex is impervious to the effects of normal aging. This observation has been shown by studies imaging aging animal models (Moreno et al., 2007; Scott A Small et al., 2004), who are unconfounded by AD, that have focused on the hippocampal formation. But beyond the hippocampal formation, these previous studies did not perform a more comprehensive comparison of the entorhinal cortex to the multiple regions of the cortical mantle.

The results from the current study imply that normal aging targets our cognitive faculties more than others. Practically, the DG, IFG, EC function and structure and the related neuropsychological tests could be used collectively for cognitive aging subject screening and intervention read-out. We also make several propositions based on our findings: 1) the spatiotemporal pattern identified in the inferior frontal gyrus is potentially important to isolate molecular or cellular causes of age-related frontal cortex dysfunction; 2) the mechanisms that account for the entorhinal's resistance to normal aging is likely linked to its vulnerability to AD.

3 Specific Aim 2: Alzheimer's disease diagnosis and normal aging evaluation based on deep learning neuroimaging

In this section, I performed a series of deep learning neuroimaging studies for diagnosis and regionality analyses. I proposed an AD diagnosis method using a 3D convolutional neural network model trained and evaluated on ~4,600 structural MRI scans and investigated a series of novel regionality analyses. The summary measures derived from the deep learning model has demonstrated potential clinical utility outperforming other AD pathology measures and biomarkers.

The AD deep learning study constitutes a general analytic framework, which I followed to perform a study on age estimation in normal aging using more than 6,000 scans. The deep learning neuroimaging approach demonstrated high classification and regression performance and also revealed regional patterns conforming to neuropathological priors. In addition, I explored the utility of deep learning on AD diagnosis and regionality analysis using PET, further demonstrating the broad utility and generalizability of the method.

3.1 Alzheimer's disease diagnosis

3.1.1 Introduction

3.1.1.1 AD diagnosis using MRI

Because of the degenerative nature of Alzheimer's disease and the current limit in therapy options, much research focuses on developing accurate diagnosis techniques. Our work applies deep learning to classify clinical diagnosis of AD from MRI, building upon recent studies that have demonstrated the usefulness of MRI in diagnosing AD and mild cognitive impairment (MCI) (overlap with the corresponding prodromal stage of AD), and in categorizing biomarkers associated with neurodegeneration in AD (Clifford R. Jack et al., 2016).

Among different brain MRI modalities, T1-weighted structural MRI is one of the most widely acquired and enjoys the additional benefit of being relatively standardized across scanners and protocols. Consequently, diagnosis algorithms based on T1w structural MRIs are appealing as a potential tool to assist in disease screening given the wide availability of research scans for model training, and the ubiquity of MRI scanners in the world potentially enabling the rapid deployment of learned models.

3.1.1.2 Deep learning

Following applications in computer vision, deep learning techniques have emerged as effective tools for analysing medical images. On standard computer vision tasks such as classification (Krizhevsky, Sutskever, & Hinton, 2012), object detection (Girshick, 2015) and semantic segmentation (Long, Shelhamer, & Darrell, 2015), deep learning algorithms based on convolutional neural network (CNN) (LeCun et al., 1990) have achieved undisputed dominance. Over the last few years, these techniques have been widely applied in image-aided medical diagnosis. Successful applications of deep learning in medical imaging include segmenting electron microscopy images (Ronneberger, Fischer, & Brox, 2015), detecting diabetic retinopathy from 2D retinal fundus photographs (Gulshan et al., 2016), and recognizing skin cancer from photographs (Esteva et al., 2017).

However, learning from 3D scans, such as MRI, presents a number of additional challenges. While the number of voxels corresponding to the 3D volume representing a single patient can be large, we still have just one label per scan, raising technical questions about how to prevent overfitting. However, since many brain disorders correspond to both focal and diffuse involvement and brain regions span 3D spaces, machine learning models capable of acting upon the whole volume are appealing. To that end, I explored the use of 3D convolutional neural networks for diagnosing Alzheimer's disease, considering a variety of techniques and including some unconventional data sources in order to learn good representations without overfitting. Although 3D CNNs have been explored in the medical imaging (Çiçek, Abdulkadir, Lienkamp, Brox, & Ronneberger, 2016; Dou et al., 2016; Milletari, Navab, & Ahmadi, 2016) for segmentations, they are relatively underutilized compared to 2D CNNs and thus best practices for deploying these models are less firmly established.

3.1.1.3 Previous works

This work was built upon several earlier works applying 3D CNNs to AD diagnosis, which were summarized below. Note that various papers address different datasets and different subsets even within the same cohort, making direct performance comparison across papers difficult. (Payan & Montana, 2015) proposed MRI-based AD diagnosis using 2265 scans from ADNI and achieved 95.39% accuracy in AD vs. CN classification. The weights of the convolution filters are learned by training an autoencoder. While their study provides promising support of the efficacy of 3D CNNs for diagnosing AD from brain MRI, it leaves open many modelling questions. For instance, they consider only shallow networks consisting of a single convolutional layer, followed by a pooling operation and one fully-connected layer. Moreover, they leave the autoencoder-derived filters fixed, optimizing only the weights of the fully-connected layer on the AD classification task. Following similar ideas, (Hosseini-Asl et al., 2018) presented a model using unsupervised autoencoding followed by (supervised) fine-tuning on a comparatively small dataset consisting of just 210 subjects (70 each for AD, MCI, CN) and showed impressive predictive performance with an area under the receiver operator characteristic (AUROC) of 0.993 and 99.3% accuracy in AD vs. CN classification. (S. Korolev, Safiullin, Belyaev, & Dodonova, 2017) apply a 3D network architecture, achieving AUROC of 0.88 with an accuracy of 79% in 50 AD and 61 CN subjects. They also attempt to interpret the network, introducing a heuristic technique for feature attribution. The method consists of generating predictions while obstructing various regions in the image to determine which regions impact the model's predictions in multiple passes. A recent large-scale study (Wegmayr, Aitharaju, & Buhmann, 2018) proposed using 3D CNNs directly, training weights from scratch (no unsupervised pre-training) and achieved 86% accuracy in AD/CN classification in a merged ADNI+AIBL dataset consisting of 6618 scans from CN subjects and 4476 scans from AD subjects. These studies demonstrate the promise of modern CNN architectures for extracting patterns from brain MRI.

Despite generating accurate predictions, deep learning has long been described as a black-box. Studies have been proposed to *interpret* or *explain* the classifications produced by various deep learning techniques. A recent study (Yang, Rangarajan, & Ranka, 2018) utilized multiple methods to generate visual explanations for AD classification despite relatively low classification performance (0.863 AUROC, 76.6% accuracy) on a small dataset consisting of 47 AD and 56 CN subjects. Ablation tests constitute one classic method for probing a predictive problem to assess the usefulness of various features by dropping the features from consideration. Our study complements the saliency map based analyses with ablation tests to assess the predictiveness of each region to AD.

Most previous studies on AD classification with CNNs (Hosseini-Asl et al., 2018; S. Korolev et al., 2017) exploit the (single) baseline scan for each subject and train their models to predict the cross-sectional diagnosis labels (assessed at scan time). (Payan & Montana, 2015) utilizes multiple scans but did not explicitly address potential data leakage and disease progression problems. (Wegmayr et al., 2018) explicitly addresses the data leakage problem (they refer to it as subject duplication), ensuring non-overlapping subjects in training and test sets, however, the disease progression is not explicitly discussed while stating several scans of an individual subject typically have the same disease label, which could be potentially problematic especially for their three-way classification task including MCI. Another difference from this previous study is that we opted to only include multiple scans from different sessions rather than within-session duplicate scans to leverage the data richness vs. data redundancy.

3.1.1.4 Contributions

In this study, we particularly focus on two specific aspects: (i) incorporating longitudinal scans as an unconventional data source, and (ii) a thorough investigation aimed at localizing the most predictive regions. Generally, data augmentation helps to prevent models from overfitting. In typical photographic images, such transformations include random crops, translations, rotations, etc. (Krizhevsky et al., 2012). Here, instead, we enrich the data by including images captured from the same patient across multiple visits. Inclusion of longitudinal scans raises two important issues: the data leakage problem and the disease progression problem. Data leakage occurs when the training and test sets contain the different scans from the same subject, the model might make the prediction by memorizing and retrieving the label from the same person and is likely to result in over-optimistic performance. This issue is dealt with easily by ensuring that the partition into train/validation/test splits takes place at the level of the individual patients instead of at the level of the scan. The disease progression problem relates to the fact that the disease status of subjects might change in follow-ups, and the cross-sectional diagnosis labels for a certain scan might be different from the baseline label, this is especially important in prodromal disease status, e.g. MCI in AD pathology.

Our model achieved high classification performance in AD classification in a large dataset consisting of 2817 scans/sessions from CN subjects and 1874 scans/sessions from AD subjects. Moreover, we found that by applying model trained on AD vs. CN classification to baseline scans of MCI subjects, we could accurately predict progression from MCI to AD. Furthermore, I studied the neuroanatomical underpinning in AD classification with a series of novel regional analyses following the regional vulnerability idea (Scott A. Small, 2014). The analyses pinpointed hippocampal formation as a most predictive driver for our deep learning-based AD diagnosis model, which further affirms the prominence of hippocampal formation in AD classification. I explored regional significance in a number of ways. In one approach, I generated a 3D saliency map *post hoc* without any changes to the model or input data, presenting the whole brain volume to the classifier and inferring regions that contribute most to the classification using a 3D class activation mapping (CAM) technique (Zhou, Khosla, Lapedriza, Oliva, & Torralba, 2016). Gradient-weighted CAM (grad-CAM) (Selvaraju et al., 2017) generalize the CAM to broader CNN families by flowing the gradients of the target label into the last convolutional layer. In this study, I utilized 3D grad-CAM to determine which specific 3D regions most indicate a prediction of AD. I also explored two “ablation”-based methods that explicitly focus on specific brain regions. The first method consists of training models to predict AD using 2D MRI slices (in each of the three coordinate planes), and evaluate the pattern of the capabilities of slices differentiating AD.

The second method, more informed of neuroanatomy, consists of masking specific regions using masks generated from a sub-population with segmentation, training on the masked regions, and comparing the classification accuracy.

The CAM method reveals a preponderance of activation overlap containing the left anterior hippocampal formation (HF). Evaluation on 2D MRI slices demonstrates the importance of slices covering the HF in the classification of AD using deep 2D CNN model. And evaluation on isolated brain lobes also demonstrates the importance of temporal lobe, which contains the HF. These findings have implications for both the interpretability of CNNs used in image-based disease diagnosis and also the prospective MRI acquisition protocols targeting AD diagnosis. Even for highly complex and nonlinear models, regionality and the underlying pathology revealed still manifests importance.

Importantly, it should be noted that our proposed 3D CNN model and regional analyses constitute a highly general framework that can potentially be applied to other brain disorders and imaging modalities.

3.1.2 Methods

3.1.2.1 Data

The dataset used in this study is from the Alzheimer's Disease Neuroimaging Initiative[5]. The details about the MRI data acquisition can be found in ADNI website[6]. The T1-weighted structural

---

[5] http://adni.loni.usc.edu/

[6] http://adni.loni.usc.edu/methods/documents/mri-protocols/

MRI scans were pre-processed with the standard Mayo Clinic pipeline[7]. The AD diagnosis was based on clinical evaluations. The MRI and diagnosis data were queried and accessed at August 2017. The diagnosis includes AD, MCI, and cognitively normal (CN). The dataset used to generate the regional masks includes 382 scans from unique elderly subjects in ADNI-2 and 1113 scans from unique young subjects in Human Connectome Project (HCP)[8] (David C. Van Essen et al., 2013). The subjects cover age 25-90, and clinical diagnosis of young normal control (N = 1113), elderly normal control (N = 120), MCI (N = 138) and AD (N = 124).

For the experiment of AD vs. CN classification, we sought to include as many MRI sessions as possible that correspond to AD or CN diagnosis at scan time. Specifically, I included the baseline and follow-up scans of patients diagnosed as AD at baseline, the baseline and follow-up scans of subjects diagnosed as cognitively normal at baseline before the conversion to AD or MCI if ever happened, and the after-conversion follow-up scans of subjects who were CN or MCI at baseline but later progressed to AD. In total, I included 4691 scans (2817 CN, 1874 AD) from 1189 subjects under these criteria. This experiment setup basically sets an upper limit to the amount of data for cross-sectional AD vs. CN classification in ADNI cohort. The inclusion of scans after conversion also helps enrich the samples around the “classification boundary” as will be more thoroughly discussed in the 3.1.2.3 “Inclusion of longitudinal scans” section.

I used 8⁄10 of the subjects as training set consisted of 3709 scans (2240 CN, 1469 AD) from 952 subjects, 1⁄10 as validation set consisted of 456 scans (279 CN, 177 AD) from 119 subjects, and 1⁄10 as test set consisted of 526 scans (298 CN, 228 AD) from 118 subjects. As

---

[7] http://adni.loni.usc.edu/methods/mri-analysis/mri-pre-processing/

[8] https://humanconnectome.org/

discussed in the Introduction section in this chapter, the training, validation and test sets split was partitioned at subject level through stratified random sampling on baseline diagnosis labels so that the groups have non-overlapping subjects and approximately even distribution of baseline diagnosis labels. The final model is selected as the one that has the highest validation accuracy (i.e. classification accuracy in validation set).

For the experiment of MCI progression prediction on baseline scans, I trained another model only using subjects whose *baseline* diagnosis are cognitively normal or AD to prevent data leakage. I included scans from 796 subjects under this criterion. Similarly, 2918 scans (1943 CN, 975 AD) from 626 subjects were used as training set, 382 scans (251 CN, 131 AD) from 80 subjects were used as validation set, 325 scans 229 CN, 96 AD) from 80 subjects were used as test set. I used the same neural network training setup. I included 318 MCI stable subjects and 311 MCI progression subjects for MCI progression prediction. The MCI stable subject are those who remained MCI during a follow-up period of at least 3 years from the initial visit. The MCI progression subjects progressed to AD at follow-up visits, among which 256 subjects progressed to AD within 3 years.

We further evaluated our model on an independent AD dataset MIRIAD[9] (Malone et al., 2013). MIRIAD includes 455 MRI scans from 46 AD subjects, and 243 MRI scans from 23 cognitively normal subjects.

---

[9] https://www.ucl.ac.uk/drc/research/methods/minimal-interval-resonance-imaging-alzheimers-disease-miriad

We also derived AUROC using bilateral hippocampal volume normalized by intracranial volume generated from FreeSurfer (Bruce Fischl et al., 2002) on the subject-level evaluations for comparison.

3.1.2.2 Preprocessing

Basic pre-processing steps include nonparametric nonuniform intensity normalization (N3) based bias field correction (Sled, Zijdenbos, & Evans, 1998), brain extraction using FreeSurfer (Ségonne et al., 2004), and 12 degree of freedom affine registration (using FSL FLIRT (Jenkinson, Bannister, Brady, & Smith, 2002) with normalized mutual information cost function) to the 1*mm* isotropic MNI152 brain template. The dimension of the 3D volume is 182 × 218 × 182 (LR×AP×SI).

Bias field correction is generally robust, fast, and based on physics models which act as a strong prior (Sled et al., 1998). There are brain extraction methods using deep learning techniques, but there is not one that is well-validated and widely-available. Skull-stripping using FreeSurfer provides consistently high-quality brain extraction.

The registration is to ensure same orientation and roughly same spatial correspondence of different images. Although there are techniques such as spatial transformer network (Jaderberg, Simonyan, & Zisserman, 2015) to learn transformation within the network, it would involve more parameters to learn and add burden to the data. All the brain extracted, affine-transformed images were checked by a well-trained reviewer with visual inspection. Scans having severe MRI artefacts, brain extraction failure or poor registration were excluded.

3.1.2.3 Inclusion of longitudinal data

Commonly, computer vision practitioners synthetically augment their datasets, applying random transformations to existing training images. However, unlike natural images or certain kinds of medical image, where objects of interest might vary in location and rotational orientation, MRI images of brains are approximately at the same position through registration, with the brain regions roughly aligned. Thus, in this setting, learning rotational and translational invariances is not well motivated.

There is another form of data augmentation or more precisely “data source” specific to medical imaging applications. For longitudinal studies, test-retest studies and just ordinary studies, there might be multiple scans per subject. By including time as a factor in subject identification, we can increase the amount of data. In a sense, by including these data sources, we are seeking natural forms of data augmentation. The corresponding “transformations” would be normal aging, disease progression or both (longitudinal scans with a significant interval between scans), subject re-positioning (scans acquired at different sessions and within a short period of time) and subject motion (scans acquired at the same session). The variability present in the scans or the data coverage in the data space decreases in this order. The illustration of the general idea is shown in Figure 3-1. As discussed in the Introduction section in this chapter about disease progression, special attention is required for the first kind, where the different time points of the same individual might be at different health or disease stages. Moreover, those scans, lying on the verge of different diagnosis, constitute informative cases for the classification.

In this study, I opted for using scans from different sessions, which already provides a significant increase in the amount of data: from 796 baseline scans to 4691 scans. And scans from the same scanning session present very low variability.

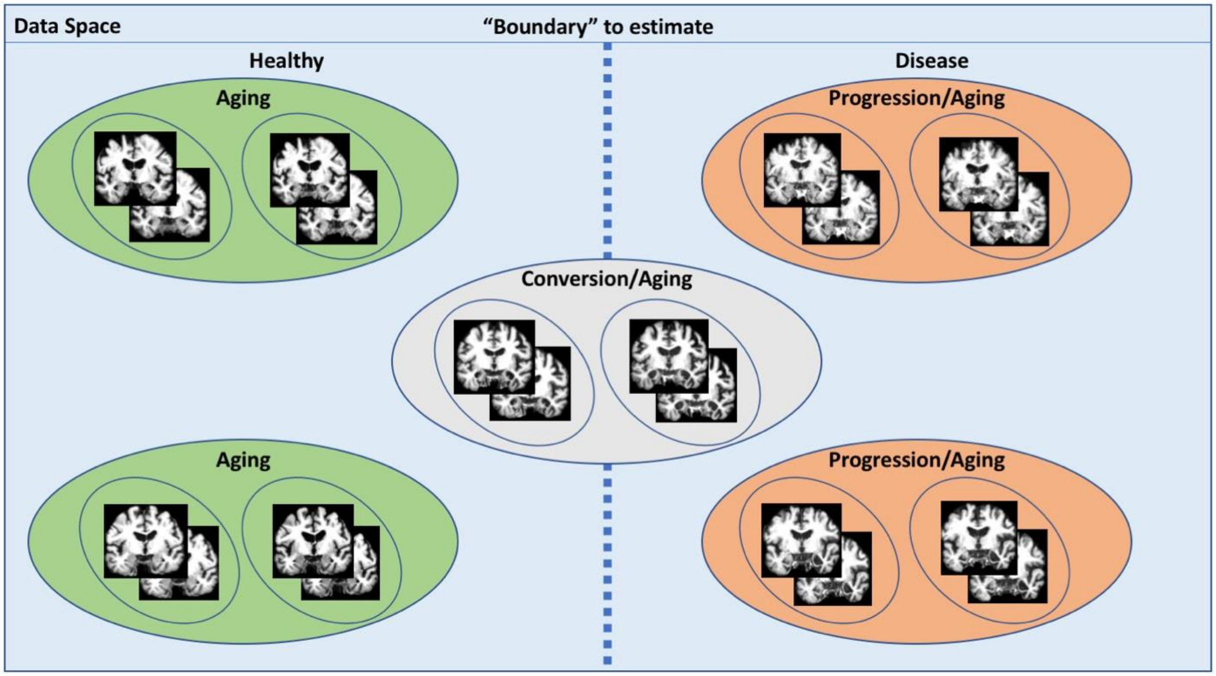


Figure 3-1. Illustration of data augmentation or inclusion of longitudinal scans specific to longitudinal neuroimaging studies for machine learning models. The whole plane is a simplified representation of the data space. Each large circle indicates one individual subject, each small circle indicates one MRI session. Each coronal slice of MRI scan represents one scan. The objective of the deep learning algorithm is to find the "boundary" (dashed line) that best differentiates cognitively normal subjects and AD patients. Enriching our data by using longitudinal scans from subjects helps to increase the data coverage from the small circle to the large circle.

3.1.2.4 Convolutional neural network

I used a general CNN architecture similar to the VGG classification architecture (Simonyan & Zisserman, 2014) with multiple interleaved convolutional blocks and max pooling layers and increasing number of features along the depth. The main differences include having just one fully-connected layer to significantly reduce the number of parameters and replacing 2D operations with 3D operations. For convolutional layers, I used a convolutional kernel size of 3 × 3 × 3, a batch size of five, rectified linear unit (ReLU) as the activation functions. The output from the last convolutional layer is flattened and fed into a fully-connected (FC) layer with sigmoid as the activation. I used batch normalization (BN) before the activation function.

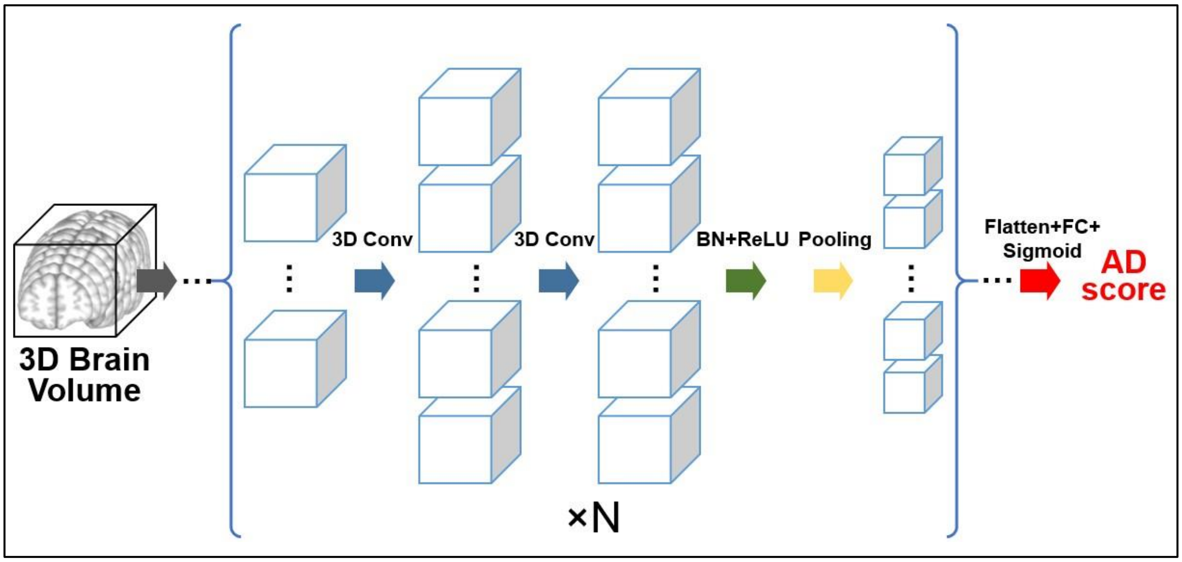


Figure 3-2. The convolution neural network architecture for AD classification. The inputs are 3D brain volumes. Each cubic represents one 3D feature map, the size reflects the spatial dimension of the feature map, and the number reflects the number of feature maps (channel dimension). The blue arrows are 3D convolutional operations, the green arrow represents batch normalization (BN) followed by rectified linear unit (ReLU), the yellow arrow denotes the max pooling operation. The basic unit enclosed in the bracket is repeated $N = 5$ times with increasing number of features and decreasing spatial dimension. The final convolutional output is flattened and fed into one fully-connected (FC) layer with sigmoid activation function (red arrow), generating the final AD score, a value between 0 and 1 reflecting the likelihood of diagnosis.

The algorithm was optimized using Adam method with cross-entropy loss function. The initial learning rate was tuned in the range from 1e-4 to 1e-6 including [1e-4, 5e-5, 2e-5, 1e-5, 5e-6, 2e-6, 1e-6] and was set at 2e-5. The algorithm was implemented using Keras and TensorFlow. As early stopping criteria, I set the patience parameter on validation accuracy to 10 epochs. I included weight *l2* regularization to prevent overfitting with a factor of 1.0. An illustration of the framework is shown in Figure 3-2. In this study, I used five (N in Figure 3-2) stages. The feature dimension of the first layer is 16 and increases by a factor of two in each subsequent stage.

3.1.2.5 Application to MCI progression prediction

The classification was trained on AD vs. CN, which presents the largest neuroanatomical contrast in the AD spectrum. Whether the model learned using the two ends of the spectrum could inform the differentiation of patients in the middle of the spectrum is critical. The model was directly applied on the baseline scans of patients diagnosed with MCI since MCI patients are not part of the training dataset of the model. AUROC was used to evaluate the prediction performance.

3.1.2.6 Class activation map

Class activation map (CAM) original proposed by (Zhou et al., 2016), extended and generalized in gradient-weighted CAM (grad-CAM) (Selvaraju et al., 2017), has been used in medical image analysis field to inform the “attention” of the 2D classification (Feng, Yang, Laine, & Angelini, 2017). In this study, I generated 3D grad-CAMs to visualize the predictive contribution of brain regions to the AD classification task. Importantly, since the map can be generated individually, it can be used as an individual neuroanatomical validity report without sacrificing the prediction power of whole-brain based prediction model. I generated the average class activation map for all AD patients to demonstrate the average "attention" of the algorithm.

For class activation mapping, the classification score can be formulated as follows:

$$AD_{score} = \sigma\left(\sum_{i}^{n} w_i \cdot a_i\right)$$

where σ is softmax function, w_i is the learned weight for each feature map A_i, and $a_i = \sum_{(x,y)} A_i(x, y)$. Following this formulation, the class activation map can be expressed as follows:

$$ADCAM = \sum_{i}^{n} w_i \cdot A_i$$

3.1.2.7 MRI 2D slice based classification

Besides the *post hoc* saliency map based class activation map method, I also propose ablation analyses focusing on part of the input data. I tested the classification using 2D CNN with the input being three consecutive slices as three channels. This design takes the inter-subject alignment precision into consideration (i.e. not extracting just one slice) and also ensuring relative similarity among different channels (i.e. not extracting five slices). The network architecture is the same as the 3D CNN architecture described above except that the 3D operations are all replaced with the corresponding 2D operations. The classification performance on different groups of 2D slices as the indication of predictive importance were reported.

3.1.2.8 Brain lobe based classification

Slice-based regional analysis method investigates the predictive regions of the classification from imaging perspective, as the coordinate planes are imaging planes. But each slice still represents a mixture of multiple regions located at a certain spatial level. It is more appealing to generate neuroanatomically meaningful regions and perform classification focusing on these regions separately. A probabilistic spatial distribution of different regions was derived from the affinely co-registered FreeSurfer segmentations (Bruce Fischl et al., 2002; B. Fischl et al., 2004) from 1,495 scans as detailed in the 3.1.2.1 Data section. An occurrence probability of 0.5% was used as the threshold for the lobe mask generation. The definition of lobes in FreeSurfer segmentation nomenclature is referenced in FreeSurfer[10], with the exception that temporal lobe also includes medial temporal lobe structures (hippocampus, amygdala). I focused on the lobe-level ablation study because the brain lobes are functionally and structurally distinct units, and also because finer region parcellation results in poor overlap across subjects and inevitable involvement of neighbouring regions.

3.1.3 Results

3.1.3.1 AD classification

The classification performance of AD vs. CN task is shown in Figure 3-3. I evaluated the model both on unique MRI sessions and the first scans of unique subjects. Our model achieves 0.980

---

[10] https://surfer.nmr.mgh.harvard.edu/fswiki/CorticalParcellation/

AUROC and 93.3% accuracy at scan level, and 0.990 AUROC and 96.6% accuracy at subject level. The AUROC evaluated using hippocampal volume is 0.916.

For the independent MIRIAD dataset, the ROC curves can be found in Figure 3-4, showing similar performance: 0.980 AUROC and 92.7% accuracy at subject level and 0.987 AUROC and 94.4% accuracy at scan level. The AUROC evaluated using hippocampal volume is 0.895.

The high overall classification accuracy of our model lays a solid foundation for our subsequent results investigating regional attribution. The model training process with only baseline scans under identical training settings stuck at uniform classification label.

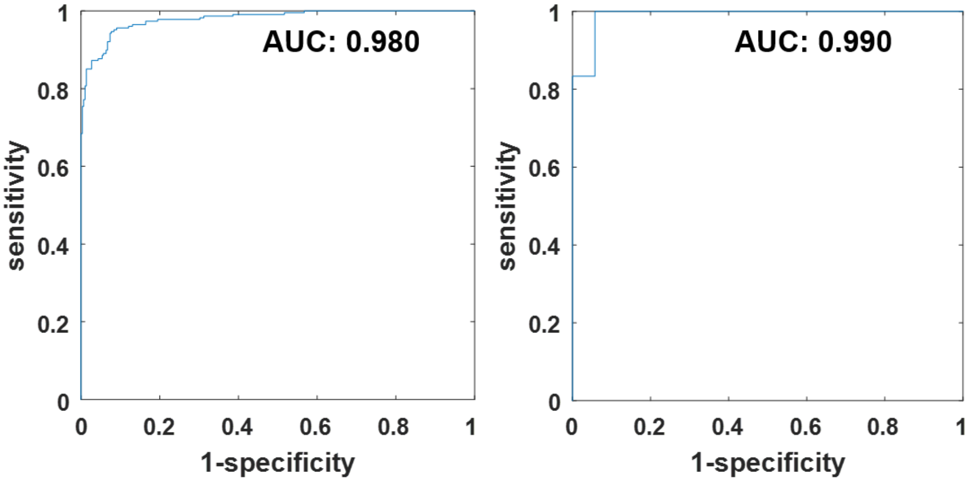


Figure 3-3. ROC curves for AD classification on the test set at (Left) scan level and (Right) subject level. The AUROCs are annotated in the figures.

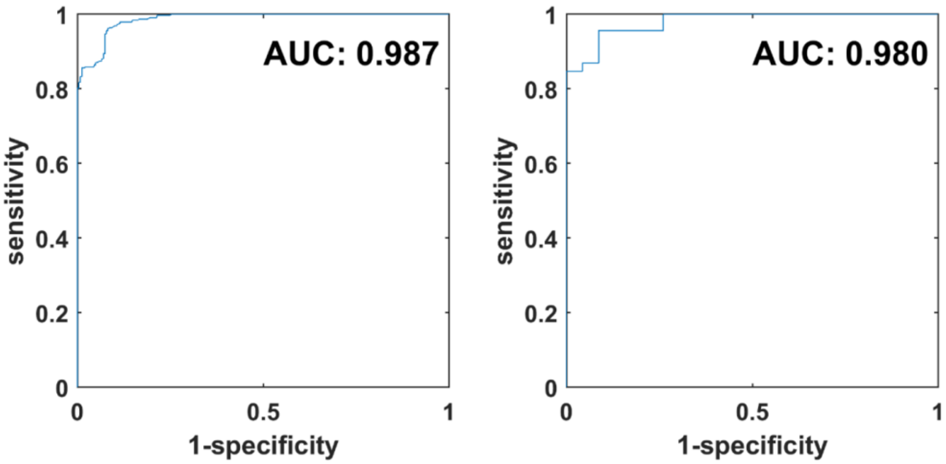


Figure 3-4. ROC curves for MRI-based AD classification in an independent dataset MIRIAD at the (Left) scan-level and (Right) subject-level. The AUROCs are annotated in the figures.

3.1.3.2 Application to MCI progression prediction

Remarkably, we found that the classifier trained exclusively on AD and CN patients can also be used *post hoc* to differentiate among those MCI patients who will and will not progress in the near term to AD. The ADNI dataset contains MCI patients whose subsequent progression or not to AD has been noted retroactively. Ideally, we might train a model exclusively on MCI patients whose subsequent progression status has been observed, directly learning to distinguish AD's prodromal stage from other causes of MCI. However, the subset of MCI patients (around 600) is not sufficient to train such a classifier but is sufficiently large to serve as an evaluation set.

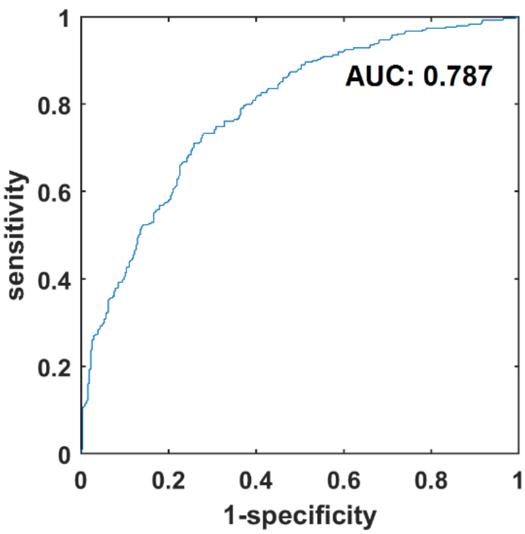


Figure 3-5. ROC curve for MRI-based MCI progression prediction, with the AUROC annotated in the figure.

To determine the usefulness of our AD vs. CN classifier for recognizing those MCI cases that will progress to AD, the imaging data of MCI patients was fed through an AD vs. CN binary classifier, interpreting a higher probability of AD as more likely to progress to AD and a higher probability of CN as less likely to progress. For this experiment, we trained the AD vs. CN model *using only baseline scans from subjects diagnosed as either AD or CN at baseline* (as detailed in the 3.1.2.1 Data section) achieving an AUROC of 0.973 on i.i.d. holdout data. Then we fed our evaluation set of MCI patients through the classifier, achieving an AUROC of 0.787 (0.808 when including only MCI patients who progressed or stayed stable within 3 years), matching state-of-the-art performance while using structural MRI data only (I. O. Korolev, Symonds, & Bozoki, 2016). In Figure 3-5, we plot the performance of our classifier as applied *post hoc* to the task of predicting MCI progression. Note that this evaluation procedure applies the CNN out-of-sample to a subset of patients that are not represented in the training set. In general, machine learning are liable to break under distribution shift and thus our performance, despite matching the previous state-of-the-art, might still have margin from the ceiling of what we might achieve given adequate data. This result supports the concept that the neuroanatomical pattern of MCI partially lies on the normal-to-AD continuum.

3.1.3.3 Regionality analyses

3.1.3.3.1 Class activation map

The average class activation map for all AD patients overlaid on the MNI152 brain template was illustrated in Figure 3-6. A 3D rendering of the iso-surface of the AD class activation map is shown in Figure 3-7. We can see from the figures that the average AD/class activation map shows large “activation” in hippocampal formation, suggesting the importance of hippocampal formation in differentiating AD in our 3D deep CNN model. Since it is not as straightforward to show scattered activations in the figure, I generated regional class activation map values to show the extent of less focalized regions both for AD (Figure 3-8) and MCI (Figure 3-9).

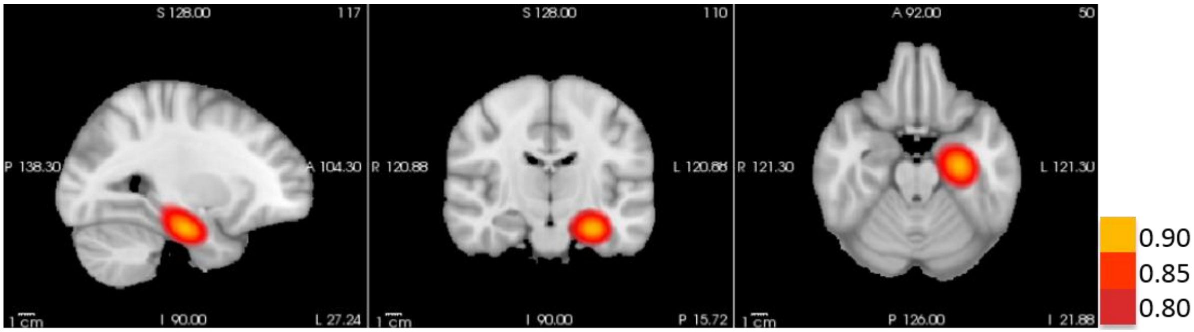


Figure 3-6. Average class activation map of AD classification overlaid on the MNI152 MRI template. The hotspot is on the hippocampal formation. The class activation map is thresholded at 0.8.

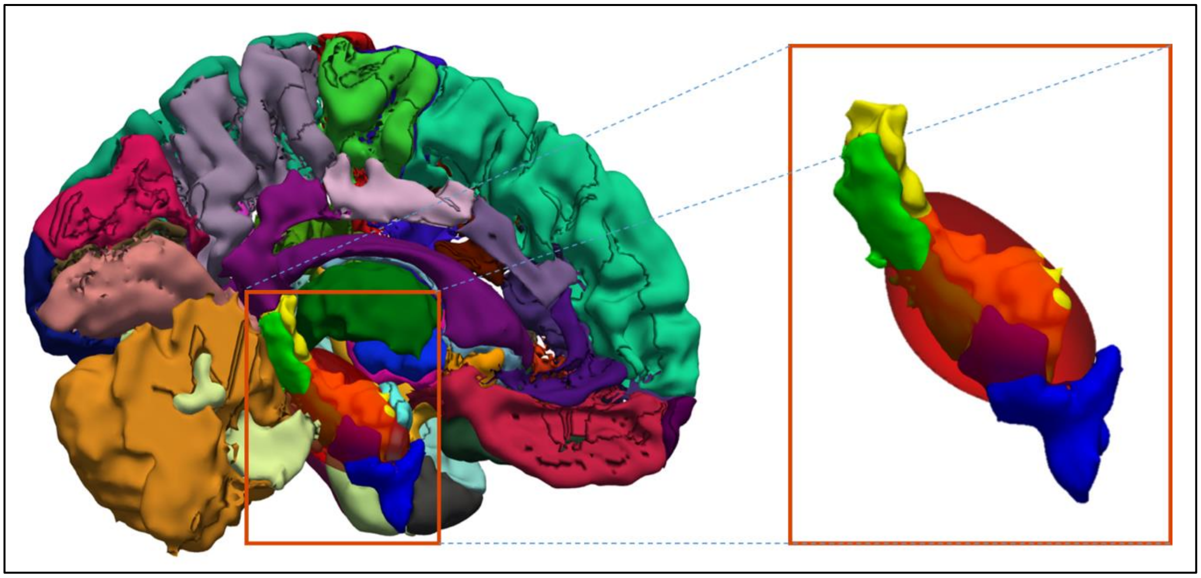


Figure 3-7. 3D rendering of the AD class activation map. The iso-surface is at the level of 0.8.

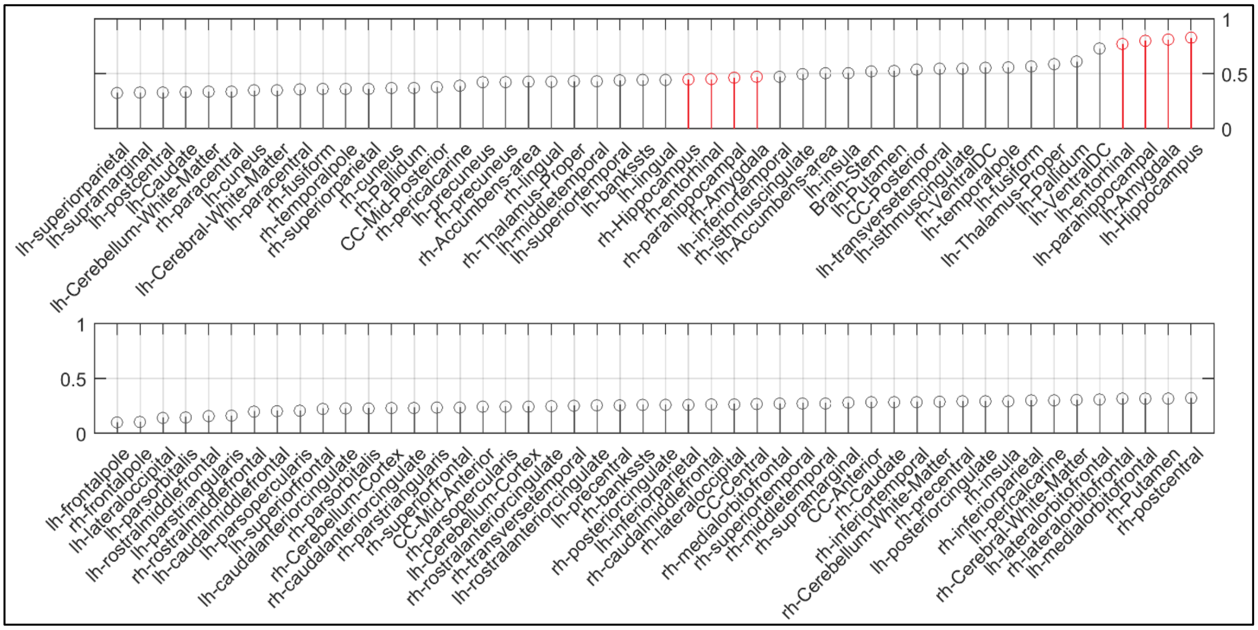


Figure 3-8. The average regional AD class activation values. The left MTL regions indicated with red color show prominence, not as prominent, the right MTL regions are also among the first quarter with around 0.5 regional values.

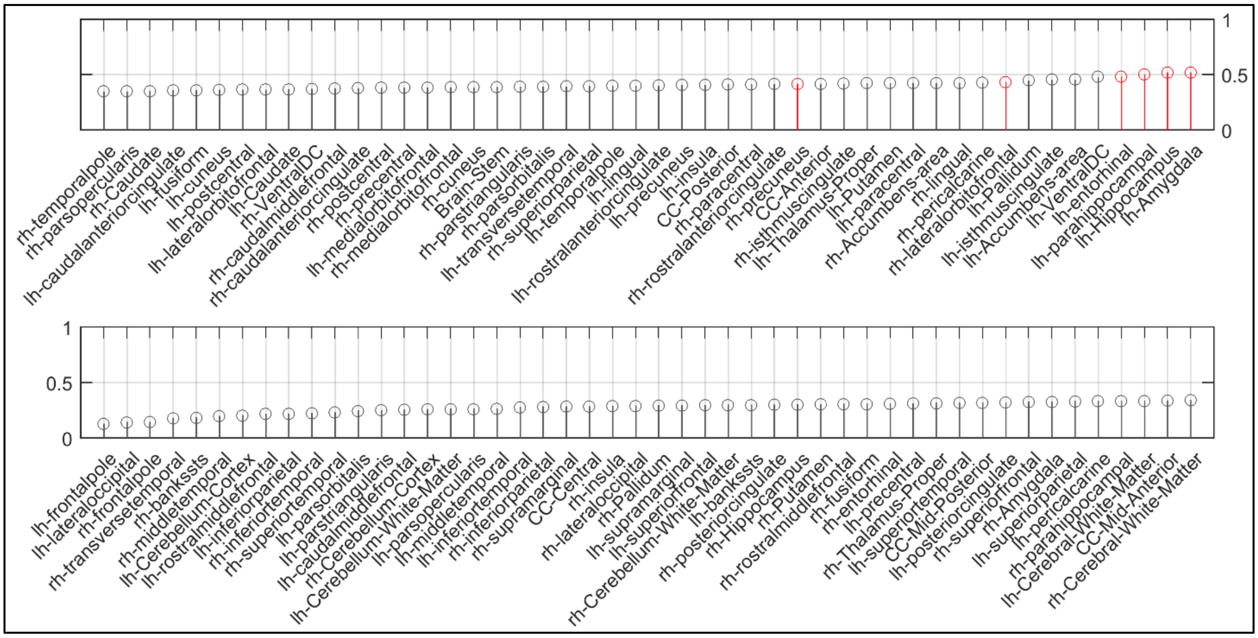


Figure 3-9. The average regional MCI class activation values. Similar to the AD class activation map, the left MTL regions indicated with red color show prominence. However, the activation map is overall more scattered with the highest regions only reaching 0.5 level. Besides the left MTL blob, there are two additional prominent blobs in parietal and frontal lobes.

3.1.3.3.2 MRI 2D slice-based classification

We also explored a slice-based classification scheme to determine which slices are most predictive of AD (Figure 3-10), running the analyses three times, using 2D slices along the sagittal, coronal, and axial dimensions. Along each dimension, those slices achieving highest classification performance include voxels belonging to hippocampal formation.

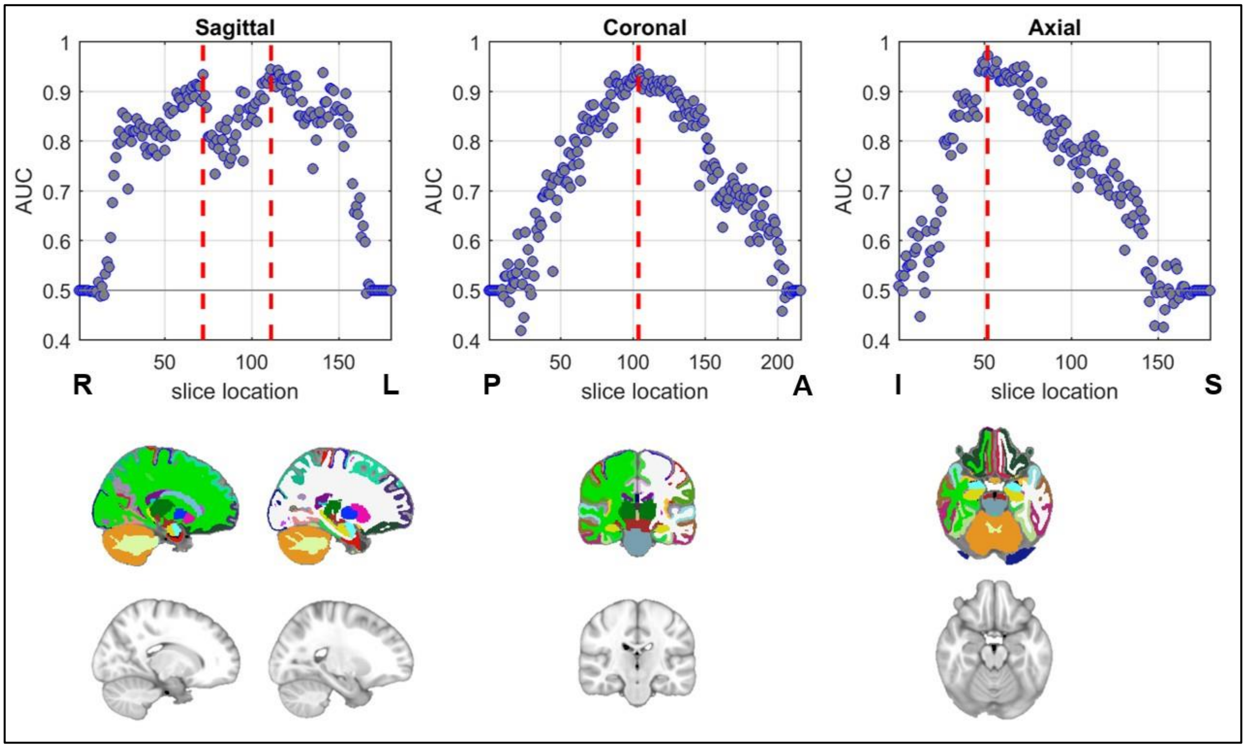


Figure 3-10. MRI 2D slice based AD classification. (Top row) The classification AUROC on the test set using 2D slices at different locations, the red lines indicate the location with highest AUROC. (Bottom row) The illustration of slices at the red lines in the top row from the MNI152 template and the corresponding regional segmentation (the colors follow the FreeSurfer color lookup table: yellow - hippocampus, red - entorhinal cortex).

3.1.3.3.3 Brain lobe based classification

In addition to the evidence for the importance of hippocampal formation provided by our 2D slice-based analysis, I also explored a more anatomically-informed method. Here, I explicitly train the model on different lobes and cerebellum, masking the others with the masks derived from a population probabilistic map. As shown in Table 3-1, the model trained on the temporal lobe, which includes hippocampal formation, achieves the highest AUROC of 0.944 and accuracy of 88%. The next most predictive lobe is the frontal lobe (0.899 AUROC and 83% accuracy).

Table 3-1. AD classification performance achieved using individual lobes and cerebellum.

Lobes	Frontal	Parietal	Temporal	Occipital	Cerebellum	Whole-brain
AUROC	0.899	0.857	0.944	0.836	0.687	0.980
Accuracy	83%	78%	88%	74%	67%	93%

3.1.4 Discussion

3.1.4.1 AD staging and dysfunction spread

In Braak staging of AD (Braak & Braak, 1991), the dysfunction represented by neurofibrillary tangles starts from transentorhinal cortex (TEC) (stage I, II) to limbic regions (stages III and IV) and spreads to neocortical at stage V and VI. Additionally, a previous fMRI study has suggested cortical spread of dysfunction originating from lateral entorhinal cortex (LEC) (Khan et al., 2014). While our findings cannot by themselves establish pathophysiological primacy, they provide evidence of structural prominence for the hippocampal formation. These results support the theory that the area circumscribing the anterior hippocampal formation is one most affected structural region in AD (Killiany et al., 2002).

3.1.4.2 Localization

Our regional analyses showed that multiple approaches of localization, including class activation maps, slice and brain lobe level ablation experiments, suggest that the hippocampal formation is the region most predictive of AD. While the interpretation of traditional statistical parametric mappings generated from mass-univariate methods (Friston et al., 1994) may be more familiar in the medical imaging community, CNNs are often considered harder to interpret, owing to complex nonlinear patterns and interactions among voxels that these models capture. As a result, they are often considered to be black-boxes, useful for pattern recognition and classification but less amenable to interpretation. The trade-offs between separability and interpretability have already been discussed in multivariate based analysis, and is becoming more obvious with the more complex architecture of deep neural networks.

However, this work highlights that through a combination of evidence produced by both heuristic saliency-map-based interpretations and rigorous slice and lobe level ablation studies, CNNs can be used not only for predictions but also to provide insights with likely neurobiological underpinnings. This work presents an important case study bridging the separability and interpretability.

While the hippocampal region does appear especially predictive of AD, we emphasize that all regions offer predictive values. Thus, in practice, for building tools to aid in the diagnosis of AD, and for predicting progression to AD among the MCI population, we recommend training models that act upon whole brain volumes. Indeed, our models acting upon whole brain volumes achieved the best AUC as compared to those acting upon any single slice or lobe.

This work relates to the multiple pathologies observations in AD (Power et al., 2018), and argues the focal neuroanatomical atrophy in hippocampal formation potentially unifies Alzheimer's disease in the presence of multiple pathologies. Since there is not a quantitative and meaningful local measure of structural information from structural MRI to do simple whole-brain voxel-based analysis or linear multivariate analysis, we deem our current approach best suited for validating this argument.

3.1.4.3 Prodromal disease classification using progressed cases

This study suggests the value of training a model on anchored disease or non-disease states with the utility of applying that model on classifying prodromal disease states; in this case MCI subjects who progress to AD. Many neurological and psychiatric diseases have known prodromal states, reflecting either mitigated or absent symptoms, or biomarkers that do not meet criteria for disease. These individuals are considered in a "high-risk" category, whereby only future classification (or failure to convert within a time frame) determines whether an individual had prodromal disease or another distinct illness. Neuroimaging may be sparse for these groups (due to difficulty in recruiting, for example, especially for rarer diseases) and such neuroimaging findings may be too subtle for traditional volumetric or segmentation based single subject analysis. By using a deep learning network enriched with data from confirmed disease states and controls, such a network may have value in screening for disease in broad populations, where the advanced disease has a distinct structural imaging signature which can easily and quickly be applied to high-risk states.

3.1.4.4 Applicability and Practicality

One advantage of our approach compared to many other neuroimaging methods is the range of MRI images onto which this technique can be feasibly applied. Given that the only pre-processing steps required are brain extraction and registration, this technique could provide a classification in a few minutes for any comparable T1-weighted image. Furthermore, the model can be retrained with more data, including data with noise or potential artefact that might prevent other standard imaging analytic techniques. It is possible that the retrained model could account for such artefacts better than traditional analytic streams.

The localization observations also have potential practical implications for acquisition. Acquisition targeting a focal field of view can facilitate diagnosis of specific disorders in the event of any time or condition constraints. And, it is important to note that the proposed method is able to generate a summary score for the structural AD-like pattern with very low computational cost at inference comparing with most of the current segmentation-based models. This is very critical in timely diagnosis and evaluation. Additionally, the computational requirement for a study this size was modest, both in cost and time, and easily adaptable into a pipeline.

3.1.4.5 Limitations

This study has several important limitations. Firstly, each MRI, even when multiple scans from the same subject are available, are treated individually. Thus our dataset contains multiple scans from the same individual and thus those individuals that have received many scans are over-represented relative to those who undergo fewer scans in model training and evaluation. Distribution of the number of scans per subject can be found in Figure 3-11. In order to ensure that the multiple scans per person do not introduce any target leaks into our learning problem, we take care to construct our training, validation and testing splits on a per-individual (vs. per-scan) basis, isolating each subject to only one split. Still it is worth noting that the most heavily scanned individuals contribute more heavily to the learned model and to the evaluation. Conceivably, if one were interested in assessing the accuracy of predictions on a per-individual level, one might re-weight the scans in the test set so that each individual contributed equally to the final score.

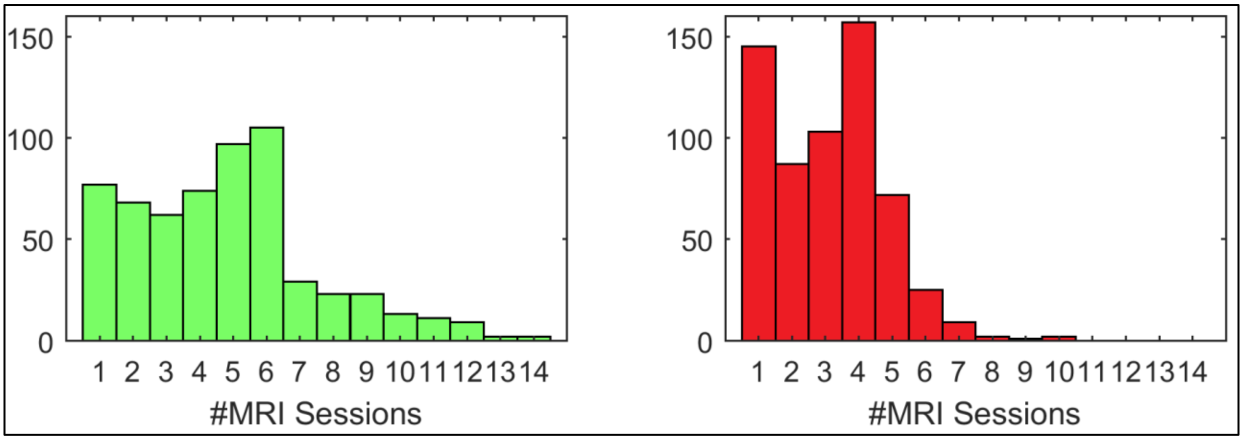


Figure 3-11. MRI sessions per subject: (left) scans of cognitively normal subjects, (right) scans of AD subjects.

Additionally, we note that our labels used in both training and evaluation are not ground truth *per se* as they are based on physicians' assessments via clinical criteria *in vivo* while true diagnosis can only be confirmed at present through neuropathology, requiring post-mortem brain biopsy. Although ante- and post-mortem concordance is generally found to be high, the availability of ground truth labels might help improve diagnosis (Beach, Monsell, Phillips, & Kukull, 2012).

The current study aims to make predictions about the diagnosis at the time of scan, which permits different diagnosis labels of multiple scans of the same subjects. Future studies with the goal of predicting longitudinal progression may define the class labels based on the most recent diagnosis label and the follow-ups.

The probabilistic spatial distribution used in brain lobe based classification experiment can be further optimized to be more reflective of general population but is not the aim and beyond the scope of this study.

3.1.4.6 Implications for the present AD biomarker system

The current AD biomarker system mainly consists of amyloid, tau, and neurodegeneration (A/T/N) (Clifford R. Jack et al., 2016), and some emerging biomarkers (Scott A. Small, Simoes-Spassov, Mayeux, & Petsko, 2017), although studies have shown the limitation of the current classification system reaching enough consistency to be used in the clinical setting (Illán-Gala et al., 2018). Machine learning could potentially bridging different biomarkers. In structural MRI, we can develop multiple models that provide structural correlates of different categories of biomarker including tau, Aβ. In this way, we are generating tau-informed or Aβ-informed neurodegenerative biomarkers. Similarly, we could derive Aβ measures from tau-PET, this way generating Aβ-informed tau biomarkers. When used for prodromal state detection, one way to interpret the measures is to extract and limit information relevant to the characteristics of an earlier phase. A preliminary analysis of generating CSF biomarker correlates from structural MRI was included in the appendix, the estimation performance still has room to improve but demonstrate the proof-of-concept.

This also has further practical implications. We can transfer the rich biomarker information from large-scale initiatives such as ADNI to other small-scale applications.

3.1.4.7 Future work

Our framework is sufficiently general that it can be easily extended to other diseases such as schizophrenia, Parkinson's disease (PD), etc. and to other MRI contrasts such as CBV, CBF or even to other imaging modalities such as PET, SPECT. One promising direction is a large-scale CBV study enabled by the recently proposed retrospective CBV technique (Feng, Hamberger, et al., 2018). Moreover, our ability to learn good representations for AD prediction could be brought to bear on other problems that might be supported only by smaller datasets. Following a number of successes in deep learning applications ranging from computer vision to natural language processing, we might employ transfer learning, fine-tuning the representations from our AD predictor, together with other sources of information such as age, gender, functional imaging measures, neuropsychological measures, CSF biomarkers, etc. to new tasks.

3.1.4.8 Conclusion

In this study, we proposed an AD diagnosis framework based on deep 3D CNN model using structural MRI, empowered with the inclusion of longitudinal scans. The proposed framework demonstrates high classification performance in AD vs. CN. In addition, we demonstrated high accuracy in MCI progression prediction applying the model trained on AD vs. CN classification to the MCI subgroup. Furthermore, through class activation map and rigorous slice and lobe-level ablation analyses, we pinpointed hippocampal formation as the most predictive regions for AD classification, affirming the prominence of hippocampal formation in AD diagnosis, and demonstrating the importance of regionality even in highly complicated deep neural network models. And importantly, the proposed classification and regional analyses methods constitute a general framework that can easily be applied to other disorders and imaging modalities.

3.2 Comparative studies to other AD pathology measures in prodromal AD detection

3.2.1 Introduction

Alzheimer’s disease (AD) progresses through a prodromal stage that presents clinically as mild cognitive impairment (MCI), before progressing to dementia (Petersen et al., 1999). MCI’s prevalence ranges between 7% to 25% of the aging population, depending on age-group and other demographics (Sachdev et al., 2015; Vos et al., 2015; Ward, Arrighi, Michels, & Cedarbaum, 2012), but only a portion of MCI patients have prodromal AD. With increased awareness and concern over AD, a growing number of MCI patients are presenting to health care providers, wanting to know whether their symptoms are caused by early stages of the AD, and if so how quickly they will progress to dementia.

Biomarkers hold promise in aiding the clinical evaluation of MCI, and biomarkers currently exist for the three core neuropathologies of AD—amyloid pathology, tau pathology, and neurodegeneration (Clifford R. Jack & Holtzman, 2013; Olsson et al., 2016). The first two can be estimated from CSF levels of Aβ and tau, or by indirect visualization using PET-sensitive radioligands. Neurodegeneration, a term currently used to encompass neuronal or synaptic loss (Khan et al., 2014), can be estimated from PET-based measures of parietal cortex metabolism, or MRI-based measurements that reflect the structural integrity of the hippocampal formation.

Large scale studies in patients with AD dementia are beginning to establish reliable diagnostic cutoffs from their parametric distributions. Applying these dementia cutoffs to dichotomize MCI patients into those who are biomarker “positive” or “negative” is one approach to determine which biomarker category (amyloid, tau, neurodegeneration) using which biomarker technology (biofluidics or neuroimaging) best classifies patients with prodromal AD. Survival analyses can then be used to determine which biomarkers best predict progression to AD dementia. This approach, however, is potentially problematic since the cutoffs are derived from patients in the dementia stage of disease (L. M. Shaw et al., 2009). Biomarkers shift over the chronically progressive course of the disease, particularly dynamic during its early stages (Anne M. Fagan et al., 2014) so that this design is prone to both type I and type II errors.

An alternative approach that can sidestep this problem is to rely on conversion to dementia information to retroactively dichotomize MCI patients into those with and without prodromal AD at baseline, and use receiver operating analyses to determine which biomarker best classifies disease. The challenge with this design is that, based on current estimates, approximately 5 years of clinical follow-up is needed in order to allot sufficient time for the majority of prodromal AD patients to clinically manifest (Mitchell & Shiri-Feshki, 2009; Visser, Kester, Jolles, & Verhey, 2006). Thanks to the ADNI, which has been acquiring biomarker data in MCI patients since 2005, this design is now possible. Here, we apply this second experimental design to ADNI data. Based on the understanding of disease pathophysiology and previous studies (Vemuri et al., 2009), we hypothesize that biomarkers of amyloid and tau pathology would outperform other biomarkers in classifying prodromal AD, but that biomarkers of neurodegeneration might outperform others in predicting time to progression.

“Deep learning” is a subset of machine learning that has proven successful in categorical image classification (LeCun et al., 2015), and when applied to MRI might prove useful in AD detection (Feng, Yang, Lipton, Small, & Provenzano, 2018). We have recently developed a deep learning algorithm that is distinguished by two assumptions: that a classifier based on individual voxels would be superior over one based on anatomical parcellation; and, that given the complex three-dimensional organization of the brain, a classifier based on 3D information would be superior over one based on 2D information. We have used this algorithm to derive from conventional MRIs individual “deep learning MRI scores” (DLMRI scores), and have found that it classifies AD with high accuracy (Feng, Yang, et al., 2018). Remarkably, the DLMRI scores are dominated by differences in the medial temporal lobe, which contains the entorhinal cortex and hippocampus. We hypothesize that, compared to other MRI-based measures of the entorhinal cortex or hippocampus, DLMRI scores will improve MRI’s clinical utility.

3.2.2 Methods

3.2.2.1 ADNI cohort

Five subsets of subjects from ADNI cohort were tested in this study: 1) ADNI MCI cohort with baseline structural MRI and *in vivo* CSF biomarker (N = 384); 2) ADNI MCI cohort with baseline structural MRI, *in vivo* CSF biomarker, FDG-PET and amyloid-PET (AV45-PET) measures (N = 248); 3) ADNI MCI longitudinal cohort with at least one follow-up measure of *in vivo* CSF biomarker and structural MRI from the same visit in addition to the baseline structural MRI and *in vivo* CSF biomarker (N = 238); 4) ADNI cohort with antemortem MRI and postmortem neuropathology (N = 44); 5) ADNI cohort with tau-PET (AV1451-PET) (N = 296). The detailed criteria in each subset are introduced in the following sections. A diagram with subject inclusion and exclusion criteria for the main dataset used in this study, subset 1 and 2, can be found in Figure 3-12.

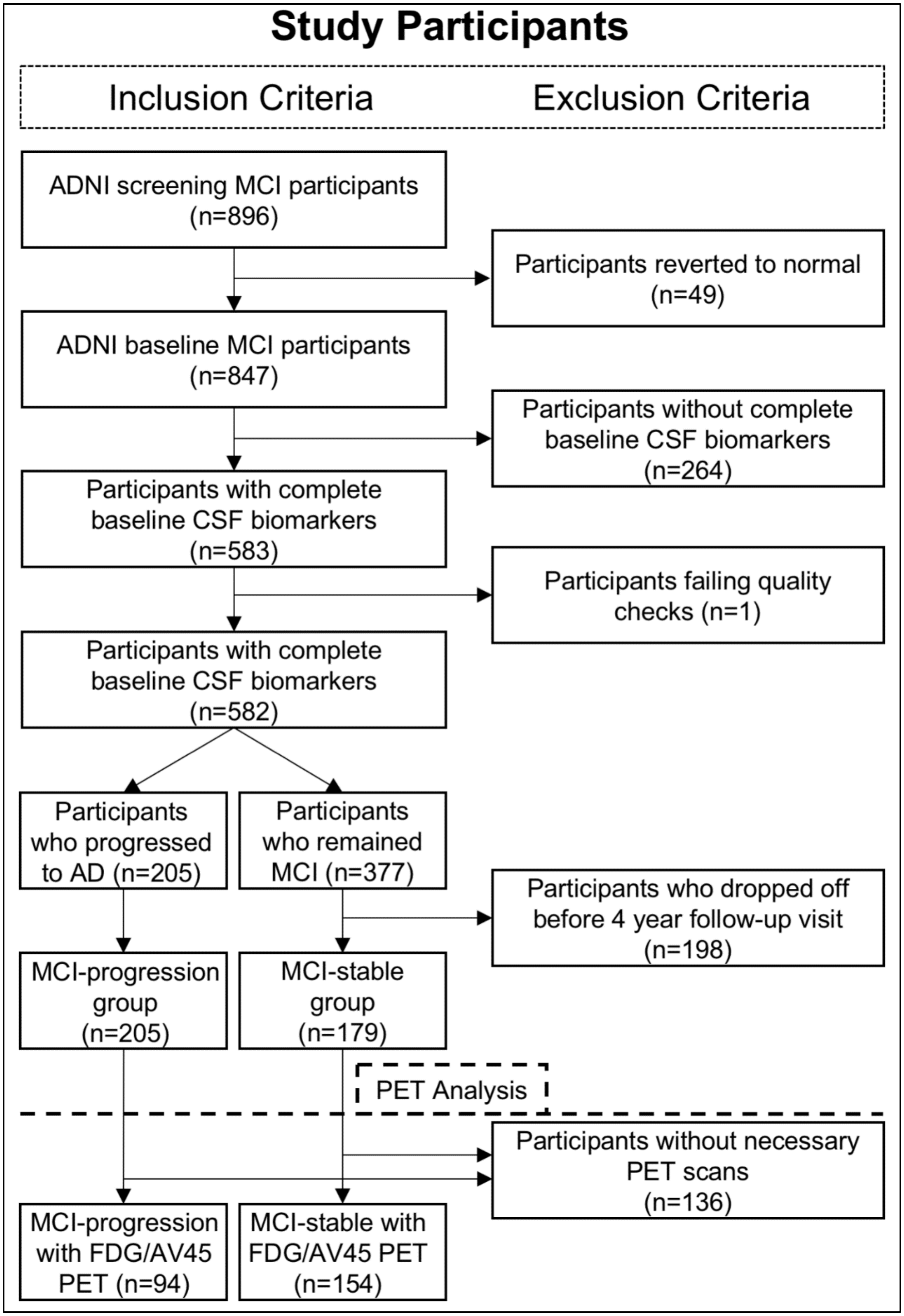


Figure 3-12. Participant selection flow-chart.

3.2.2.1.1 ADNI MCI cohort with *in vivo* CSF biomarker

Among all MCI patients with follow-up, we further confined our cohort by constraining the MCI stable criteria. The raw distributions of the follow-up lengths in the MCI-stable group, and the conversion time of MCI-progression group are shown in Figure A-4. We essentially excluded the left tail of the distribution in the MCI-stable group to decrease the proportion of potential prodromal AD incidence in the MCI-stable group.

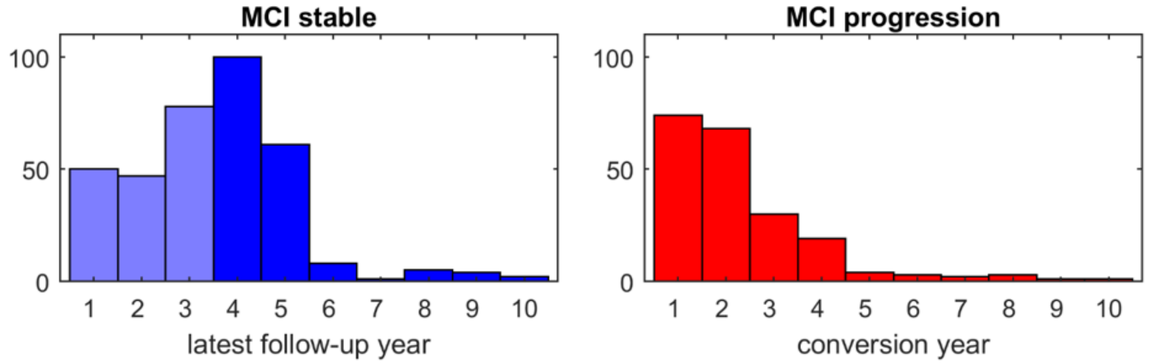
We identified 384 baseline MCI subjects who have both MRI and *in vivo* CSF biomarkers at baseline (L. M. Shaw et al., 2009). Patients in MCI-stable group stayed stable in the follow-up visits and have a follow-up period of at least 4 years. The demographic information and the baseline measures compared are summarized in Figure 3-13 (and for more details, see Figure A-4). In this constrained cohort, MCI-progression takes up 53.4% of the whole population. Within this large cohort, we generated and compared with measures including: *in vivo* CSF biomarkers, morphometric measures, and neuropsychological measures.

3.2.2.1.2 *In vivo* CSF biomarker

CSF tau levels, reflective of neurofibrillary tangle, and CSF Aβ levels, reflective of amyloid pathology, were included in the analysis (Tapiola et al., 2009). Additionally, the tau/Aβ ratio, which has been shown to best capture AD (A. M. Fagan et al., 2007), was also included. CSF was acquired at ADNI sites in accordance to the ADNI acquisition protocols and analyzed as previously described (L. M. Shaw et al., 2009). The median values provided by ADNI were used.

3.2.2.1.3 Morphometric measures

We used FreeSurfer 6.0 (Bruce Fischl et al., 2002; B. Fischl et al., 2004) to segment the same set of scans and derive regional morphometric measures. Hippocampal (HC) volume, entorhinal cortex (EC) volume, entorhinal cortex thickness were used as AD vulnerable structural measures. Hippocampal and entorhinal volume were normalized by the intra-cranial volume (ICV).



Main dataset	MCI stable N = 179	MCI progression N = 205	Total N = 384
age	71.3 ± 7	73.5 ± 7.1	72.5 ± 7.1
sex M/F (%M)	107/72 (59.8)	124/81 (60.5)	231/153 (60.2)
APOE ε4 frequency (2/1/0)	13/52/114	33/101/71	46/153/185
Last visit FU year	4.74 ± 1.2	-	-
conversion year	-	2.2 ± 1.7	-
DLMRI score	0.385 ± 0.165	0.622 ± 0.196	0.511 ± 0.217
Aβ	190.3 ± 51.3	144 ± 39.8	165.6 ± 51
tau	74.1 ± 43.0	115.3 ± 56.4	96.1 ± 54.6
tau/Aβ	0.457 ± 0.395	0.881 ± 0.532	0.684 ± 0.518
EC thickness	3.34 ± 0.4	3.03 ± 0.43	3.18 ± 0.44
EC volume (/ICV)	2.62E-03 ± 5.15E-04	2.31E-03 ± 4.95E-04	2.46E-03 ± 5.27E-04
HC volume (/ICV)	4.81E-03 ± 6.88E-04	4.14E-03 ± 6.00E-04	4.45E-03 ± 7.25E-04
MMSE	28.1 ± 1.7	26.9 ± 1.8	27.5 ± 1.8
RAVLT retention	49.6 ± 30.8	24.6 ± 29.3	36.2 ± 32.5
PET subset	**MCI stable N = 154**	**MCI progression N = 94**	**Total N = 248**
age	70.9 ± 7	72.8 ± 6.6	71.6 ± 6.9
sex M/F (%M)	88/66 (57.1)	52/42 (55.3)	140/108 (56.5)
APOE ε4 frequency (2/1/0)	11/44/99	17/50/27	28/94/126
Last visit FU year	4.43 ± 0.55	-	-
conversion year	-	1.91 ± 1.16	-
DLMRI score	0.376 ± 0.163	0.635 ± 0.190	0.474 ± 0.214
FDG	1.31 ± 0.11	1.16 ± 0.11	1.26 ± 0.13
AV45	1.15 ± 0.19	1.4 ± 0.21	1.24 ± 0.23

Figure 3-13 . Distribution and demographics of study participants. Distribution frequencies of the patients with amnestic mild cognitive impairment (MCI) at baseline, who either remained stable (MCI stable) or progressed to Alzheimer's dementia (MCI progression), organized by latest follow-up years and conversion years. The dark blue bars indicate subjects included in the study. Demographic and baseline biomarker data are listed in the table for the MCI stable and MCI progression groups.

3.2.2.1.4 Neuropsychological measures

All participants have neuropsychological measures available. For neuropsychological batteries, we tested the MMSE (Mini-Mental State Examination) score and the RAVLT (Rey Auditory Verbal Learning Test) retention score. RAVLT retention measure equals the number of delayed recalled words divided by the number of words learned in trial 5 and has been found to be one of the most sensitive to AD (Chang et al., 2010).

3.2.2.1.5 ADNI MCI cohort with PET measures

To compare DLMRI score with FDG and amyloid-PET measures, we further confined the subset into subjects who also had received baseline FDG and AV45-PET. We identified N = 248 subjects in this subset, including 154 in MCI stable group and 94 in MCI progression group. We used the composite FDG and composite AV45 scores provided by ADNI for analyses (Jagust et al., 2015). The FDG composite measure is the average FDG uptake of angular, temporal, and posterior cingulate (Susan M. Landau et al., 2011). The AV45 composite measure is the average AV45 SUVR (standard uptake value ratio) of the frontal, anterior cingulate, precuneus, and parietal cortex relative to the cerebellum (S. Landau & Jagust, 2015).

3.2.2.1.6 ADNI MCI longitudinal cohort

To investigate the longitudinal association of DLMRI score with CSF biomarkers, among the 384 subjects with baseline CSF biomarkers, we identified 238 subjects with at least one follow-up measure of CSF biomarkers and structural MRI from the same visit.

3.2.2.1.7 ADNI cohort with neuropathology

Among 64 subjects with brain autopsy data, we identified 44 and 29 subjects with an antemortem MRI scan within two and one years before death, respectively. The closer to autopsy the last MRI was acquired, the more consistency we would expect for the measures from the two different sources. We derived DLMRI scores on the last antemortem structural MRI scans in this cohort. We analyzed the association with neuropathological measures including the Braak stage for neurofibrillary tangles (Braak & Braak, 1991) and the Thal phase for amyloid plaques (Thal, Rüb, Orantes, & Braak, 2002). Additional neuropathological measures include density of neocortical neuritic plaques (CERAD score, Consortium to Establish a Registry for Alzheimer's Disease) (Mirra et al., 1991), and the Alzheimer's disease neuropathologic change (ADNC) score (Hyman et al., 2012).

3.2.2.1.8 ADNI cohort with tau-PET

ADNI began acquiring PET scan using the AV1451 radioligand, which binds neurofibrillary tangles (Marquié et al., 2015), in the late phase of ADNI2 and resumed in ADNI3. Due to the smaller number of subjects with available longitudinal data or follow-up, we performed cross-sectional analyses on these subjects with tau-PET. We used the regional tau measures generated by the UC Berkeley group provided by ADNI (Jagust et al., 2015).

3.2.2.2 Deep learning derived MRI score

The deep learning model is a 3D convolutional neural network (CNN) model, and was trained on the brain-extracted T1-weighted structural MRI scans from the ADNI cohort of AD and cognitively normal subjects (Feng, Yang, et al., 2018). The continuous output from the model is reflective of the progressive structural patterns of AD pathology. We refer to it as a deep learning derived MRI (DLMRI) score. All subsequent analyses were performed using DLMRI score. More information about the deep learning model development and validation can be found in section 3.1 and in our previous paper (Feng, Yang, et al., 2018).

3.2.2.3 Statistical analysis

3.2.2.3.1 ROC analysis

We performed receiver operating characteristic (ROC) analysis for MCI progression prediction. We analyzed the standardized residuals controlling for age, sex, and APOE ε4 frequency using linear regression. We used DeLong test (DeLong, DeLong, & Clarke-Pearson, 1988) to test for the significance of the differences in the AUROCs (area under the ROC curve) from different measures using pROC R package (Robin et al., 2011).

3.2.2.3.2 Survival analysis

Cox proportional hazards regression models were fit to examine the association between baseline measures and time to conversion to AD dementia from MCI, controlling for age, sex, and APOE ε4 frequency, using survival R package (Therneau & Grambsch, 2013). MCI-stable subjects are included in the model as censored data with the last visit as the censored point. The high-risk and low-risk survival curves were generated with 75% percentile and 25% percentile of the observed measures, respectively.

3.2.2.3.3 Longitudinal analysis

The longitudinal association between DLMRI score and CSF biomarkers was studied by examining the deviation from baseline measurements for each participant over time. The changes in either CSF biomarker or DLMRI score of all follow-up visits from baseline were used to estimate the slope β of the change in tau (Δtau), Aβ (ΔAβ), and tau/Aβ ratio (Δtau/Aβ) versus the change in DLMRI score (ΔDLMRI) for each participant using linear regression through the origin. Each participant was represented by the point based on the last follow-up visit's $\Delta\text{DLMRI}_{\text{last}}$ (x-coordinate) and the fitted change $\beta\Delta\text{DLMRI}_{\text{last}}$ (y-coordinate) of the respective measure. The last follow-up visit was used to anchor the representation of the participant in order to reflect the full follow-up. A correlation analysis was performed across participants. A linear regression model was fit across participants and illustrated. A single-subject illustration is shown in Figure 3-14.

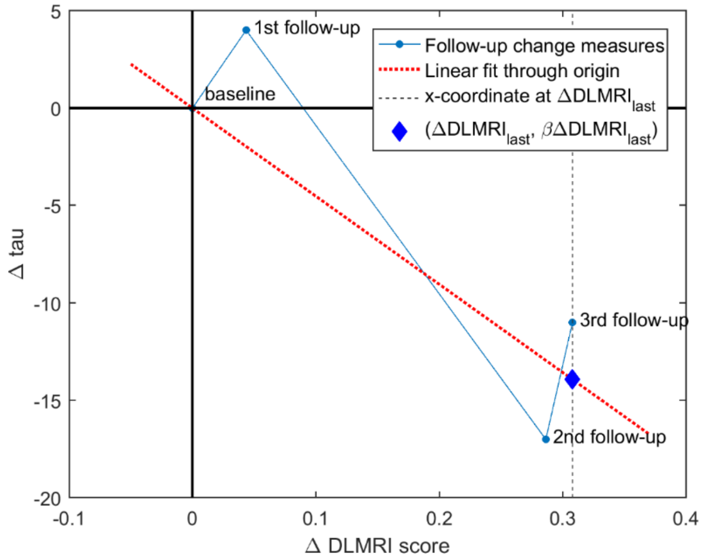


Figure 3-14. An illustration of the longitudinal measure derivation for a sample participant. The follow-up change measures from baseline were illustrated in the blue circles with the

corresponding visits annotated. Based on the follow-up change measures, a linear fit with slope β passing through the origin was estimated and illustrated as a red line. The DLMRI score at last visit (ΔDLMRIlast, x-coordinate) and the corresponding fitted change (βΔDLMRIlast, y-coordinate), indicated as the diamond in the figure, can be calculated for each participant regardless of frequency of follow-up.

And the longitudinal analysis pipeline with MMSE vs. DLMRI score as an example is shown in Figure 3-15.

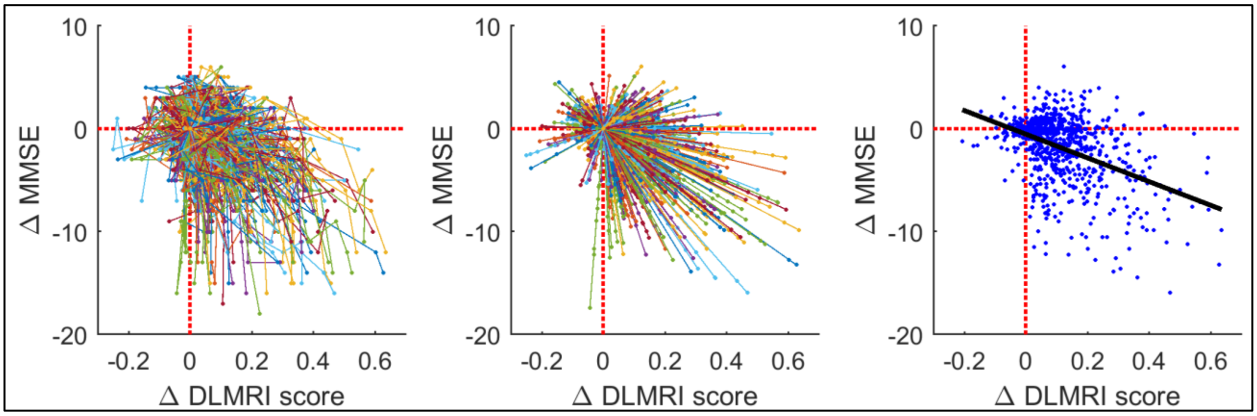


Figure 3-15. Longitudinal analysis pipeline with MMSE vs. DLMRI score as an example. (Left) The raw longitudinal data with each colored line indicating one subject and each dot representing one visit time-point; (Mid) The linear fit of each individual, each line segment starts from (0, 0) and ends at ($\Delta DLMRI_{last}$, $\beta\times\Delta DLMRI_{last}$); (Right) The linear fit of ($\Delta DLMRI_{last}$, $\beta\times\Delta DLMRI_{last}$) across subjects.

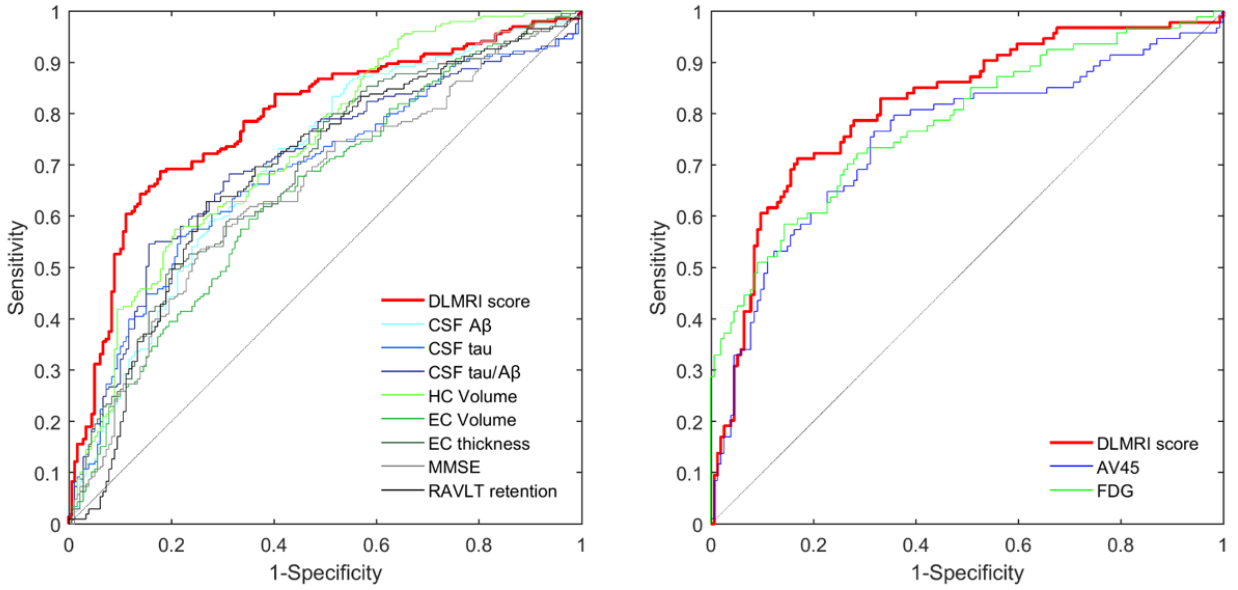
3.2.2.3.4 Correlational analysis

We performed partial correlation between baseline DLMRI score and *in vivo* CSF biomarkers, regional tau-PET measures, controlling for age, sex, and APOE ε4 frequency. As the Braak staging of neurofibrillary tangles and the Thal phase of amyloid plagues are both rank ordinal measures, we correlated the DLMRI score with the neuropathological measures using Spearman correlation.

3.2.3 Results

3.2.3.1 Classifying prodromal Alzheimer's disease

The ROC analyses revealed that the DLMRI score outperformed all other morphometric, neuropsychological, and CSF biomarker based measures at classifying MCI-stable group and MCI-progression group, with a 0.788 AUROC. AUROC (DeLong Test p values) for other measurements are Aβ: 0.0702 (0.0141), tau 0.682 (0.0161), tau/Aβ 0.703 (0.0161), hippocampal volume 0.733 (0.0484), entorhinal cortex volume 0.648 (2.01E-6), entorhinal cortex thickness 0.685 (1.71E-4), MMSE 0.648 (6.70E-5) and RAVLT retention 0.686 (2.28E-3).



	DLMRI score	Aβ	tau	tau/Aβ	HC volume	EC volume	EC thickness	MMSE	RAVLT retention	DLMRI score	AV45	FDG
AURUC	0.788	0.702	0.682	0.703	0.733	0.648	0.685	0.648	0.686	0.815	0.751	0.782
p-value	-	0.0141	2.47E-3	0.0161	0.0484	2.01E-6	1.71E-4	6.70E-5	2.28E-3	-	0.154	0.330

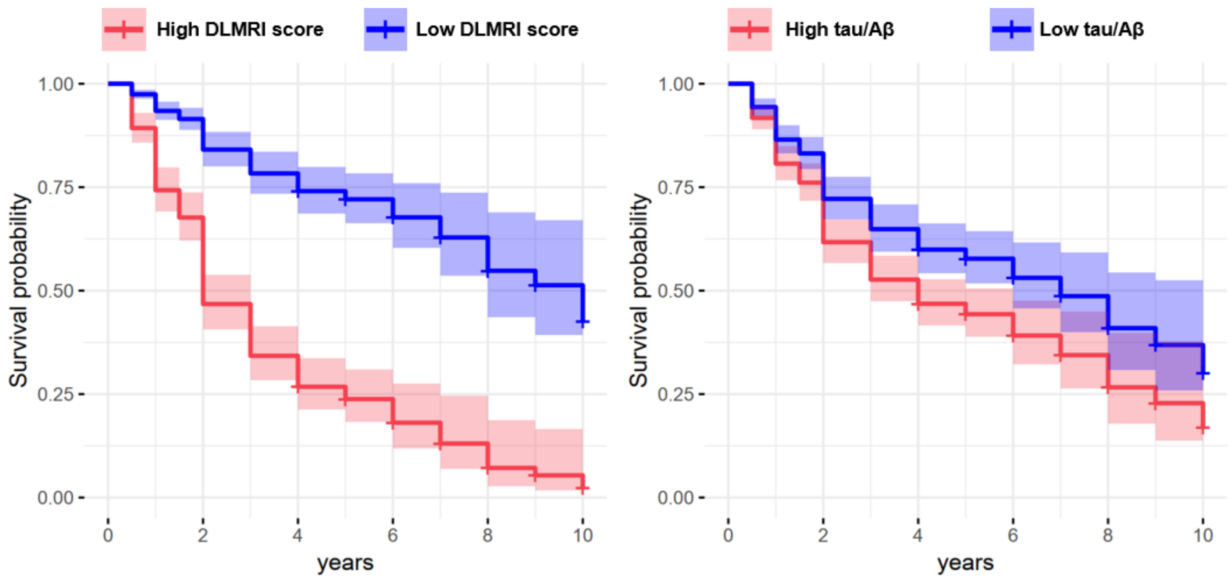
Figure 3-16. Classifying prodromal Alzheimer's disease. By comparing the 'MCI stable' to the 'MCI progression' groups, ROC curves show that the 'deep learning MRI' (DLMRI) scores were found superior in classifying prodromal Alzheimer's disease (indicated in red). DLMRI scores outperformed (left panel) CSF measures of Aβ, tau, or tau/Aβ; MRI measures of hippocampal (HC) or entorhinal cortex (EC) volume or thickness; clinical measures using the modified mental status exam (MMSE), the retention of the Rey Auditory Verbal Learning Task (RAVLT) (left panel); and, in a smaller subset (right panel), PET measures of amyloid using the AV45 radioligand or metabolism using fluorodeoxyglucose (FDG). Specific area under the curve (AUROC) values for each measure, and statistical probability values for each comparison, are shown in the table on the bottom panel.

We observe from the ROC curves in Figure 3-16 that the ROC curve of DLMRI score has higher overall sensitivity and specificity compared to the other measures for almost all levels. The differences in AUROC are statistically significant as described above and shown in Figure 3-16.

In the subset of subjects with PET measures available, we show the DLMRI score achieves better classification performance (AUROC of 0.815) than the amyloid-PET composite score (0.751) and FDG-PET composite score (0.782), though the difference is not statistically significant ($p = 0.154$ and 0.330, respectively), as shown in Figure 3-16.

3.2.3.2 Predicting progression to AD dementia

Demonstrating superior performance at identifying prodromal AD, we further compared the association of different measures with the time to conversion to AD dementia in Figure 3-17.



	log HR	**SE**	**z**	**P**	**Chi-square**	**25% quantile**	**75% quantile**
DLMRI score	4.20	0.382	11.0	4.35E-28	127.0	0.328	0.681
Aβ	-1.21E-02	1.90E-03	-6.37	1.87E-10	44.7	128.0	204.3
tau	6.79E-03	1.19E-03	5.70	1.18E-08	29.1	56.8	124.0
tau/Aβ	0.616	0.114	5.41	6.29E-08	25.0	0.285	0.924
HC volume	-1.07E+03	1.22E+02	-8.80	1.35E-18	81.5	3.95E-03	4.92E-03
EC volume	-8.64E+02	1.44E+02	-6.02	1.75E-09	37.3	2.11E-03	2.82E-03
EC thickness	-1.14	0.154	-7.42	1.21E-13	51.5	2.90	3.52
MMSE	-0.222	3.88E-02	-5.72	1.07E-08	32.2	26	29
RAVLT retention	-1.79E-02	2.60E-03	-6.88	6.12E-12	52.7	0	63.64
DLMRI score	5.02	0.554	9.05	1.40E-19	86.4	0.301	0.644
AV45	3.41	0.479	7.12	1.04E-12	48.2	1.03	1.43
FDG	-8.77	0.963	-9.11	8.14E-20	86.9	1.18	1.34

Figure 3-17. Predicting progression to Alzheimer's Dementia. Survival analyses were performed comparing the DLMRI scores to other measures, and example curves illustrate that the DLMRI score (left panel) outperformed the CSF measure of the tau/Aβ ratio (right panel). The high risk (indicated by red) and low risk (indicated by blue) curves were fitted from 75% and 25% percentile of the measures respectively. The shaded area indicates the 95% confidence interval.

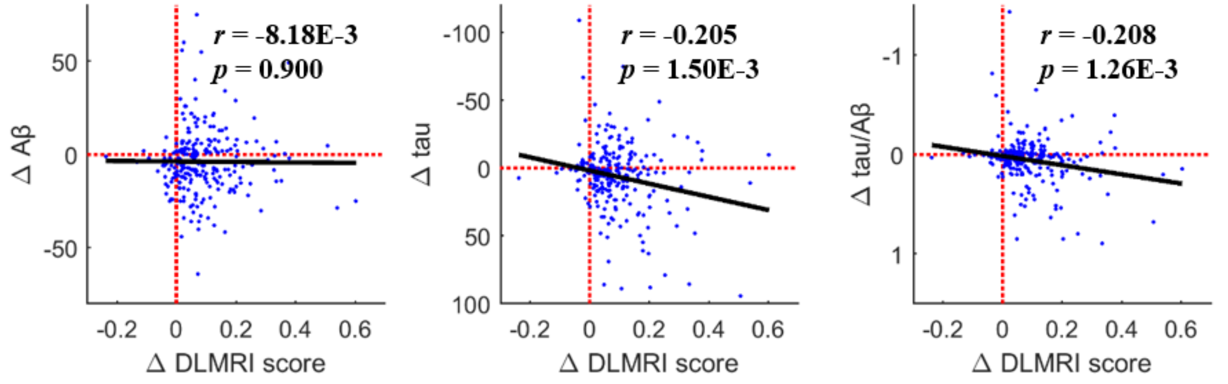
The DLMRI scores outperformed CSF Aβ, tau, or tau/Aβ, HC volume, EC volume, EC thickness, MMSE, RAVLT retention, amyloid-PET measure, and is comparable to FDG-PET measure, as shown in the table on the bottom panel.

The patterns are similar to the classification results in the previous section. DLMRI score shows better prediction capability (|z| = 11.0, p = 4.35E-28) than structural morphometric measures (HC volume |z| = 8.80, p = 1.35E-18, EC volume |z| = 6.02, p = 1.75E-09, EC thickness |z| = 7.42, p = 1.21E-13) neuropsychological measures (MMSE |z| = 5.72, p = 1.07E-08, RAVLT retention |z| = 6.88, p = 6.12E-12), CSF biomarkers (Aβ |z| = 6.37, p = 1.87E-10, tau |z| = 5.70, p = 1.18E-08, tau/Aβ |z| = 5.41, p = 6.29E-08), and amyloid-PET composite scores (DLMRI PET cohort |z| = 9.05, p = 1.40E-19, AV45 |z| = 7.12, p = 1.04E-12); and comparable to the predictive performance of FDG-PET (|z| = 9.11, p = 8.14E-20). We illustrate the survival curves of both high and low DLMRI scores and tau/Aβ ratios in Figure 3-17, at 75% and 25% percentile of the baseline measures.

3.2.3.3 Correlations with AD pathology

Besides comparative studies with other measures in the prodromal AD classification and time to conversion prediction, we also studied the association of DLMRI score with reported AD pathology measures. It was important to determine whether the DLMRI score is linked to amyloid or tau pathology. Cross-sectionally, DLMRI score is significantly correlated with both Aβ level (r = -0.190, p = 1.86E-4), tau level (r = 0.225, p = 9.00E-6), and the tau/Aβ ratio (r = 0.209, p = 3.90E-5), as also visualized in Figure A-5. The correlation with tau is marginally stronger than Aβ. Longitudinally, as shown in Figure 3-18, the DLMRI change is significantly associated with the longitudinal change in tau levels (r = -0.205, p = 1.50E-3) and tau/Aβ ratio (r = -0.208, p = 1.26E-3), but not with Aβ (r = -8.18E-3, p = 0.900).

And we further studied the association with the neuropathological measures, which are the most definitive way of characterizing AD, though with various time lag from the clinical visits in the lifetime. We show DLMRI score is more associated with Braak staging of neurofibrillary tangles than the Thal phase of amyloid plagues, as shown in Figure 3-18 and also visualized in in Figure A-6.



MRI-autopsy interval	1 year (N = 29)		2 years (N = 44)	
	corr-coef	p-value	corr-coef	p-value
Braak stage (neurofibrillary tangles)	0.402	0.0305	0.397	7.70E-3
Thal phase (amyloid plaques)	0.239	0.2119	0.196	0.2030

Figure 3-18. The "deep learning MRI" score correlates with tau pathology. The scatter plots illustrate the relationship between changes over time in the DLMRI scores vs. changes in Aβ (left panel), changes in tau (middle panel) and changes in tau/Aβ (right panel). Each data point indicates one participant's change of last DLMRI score from baseline ($\Delta DLMRI_{last}$), plotted against their change in biomarkers measures. The black solid lines are the linear fits across participants, showing that changes in the DLMRI score is most strongly correlated with changes in tau over

time. The table on the bottom panel lists the correlations between antemortem DLMRI scores to postmortem-derived Braak stage of neurofibrillary tangles and the Thal phase of amyloid plaques, with an MRI-autopsy interval of either 1 year or 2 years, showing that DLMRI scores are most strongly correlated with tau pathology.

Overall, we demonstrate DLMRI score strongly linked to global levels of both ante- and postmortem tau measurements. Thus, we further analyzed the regionality of the link with tau levels by looking into a tau-PET dataset, and show strongest correlations in MTL, particularly entorhinal cortex (r = 0.449, p = 1.66E-15). The full list of analysis results can be found in Table 3-2 for cortical regions and Table 3-3 for subcortical regions.

Table 3-2. Partial-correlation statistics between cortical regional tau-PET measures and DLMRI score. The partial correlations control for age, gender, and APOE ε4 frequency. The regions are sorted in the order of correlation coefficients.

cortical regions	**corr**	**p-value**	**cortical regions (cont.)**	**corr**	**p-value**
entorhinal	0.449	1.66E-15	insula	0.347	1.79E-09
inferiortemporal	0.440	7.45E-15	lateralorbitofrontal	0.346	2.04E-09
fusiform	0.428	4.16E-14	precentral	0.337	5.41E-09
temporalpole	0.427	5.16E-14	medialorbitofrontal	0.324	2.39E-08
parahippocampal	0.422	1.01E-13	lateraloccipital	0.321	3.06E-08
middletemporal	0.420	1.49E-13	rostralmiddlefrontal	0.318	4.25E-08
inferiorparietal	0.416	2.63E-13	lingual	0.317	4.65E-08
precuneus	0.416	2.66E-13	cuneus	0.317	4.79E-08
isthmuscingulate	0.416	2.77E-13	parstriangularis	0.307	1.30E-07
bankssts	0.415	3.03E-13	paracentral	0.296	3.83E-07
caudalmiddlefrontal	0.410	6.01E-13	pericalcarine	0.293	4.89E-07

superiorparietal	0.389	1.06E-11	parsorbitalis	0.291	5.83E-07
posteriorcingulate	0.388	1.17E-11	caudalanteriorcingulate	0.277	2.06E-06
superiorfrontal	0.380	3.48E-11	postcentral	0.271	3.63E-06
superiortemporal	0.379	3.79E-11	rostralanteriorcingulate	0.270	3.75E-06
supramarginal	0.377	5.33E-11	frontalpole	0.222	1.57E-04
parsopercularis	0.351	1.19E-09	transversetemporal	0.193	1.09E-03

Table 3-3. Partial-correlation statistics between subcortical regional tau-PET measures and DLMRI score. The partial correlations control for age, gender, and APOE ε4 frequency. The regions are sorted in the order of correlation coefficients.

Subcortical regions	**corr-coef**	**p-value**
amygdala	0.454	7.82E-16
putamen	0.336	6.37E-09
hippocampus	0.272	3.34E-06
pallidum	0.270	4.03E-06
ventral DC	0.121	0.0420
caudate	0.0923	0.121
thalamus-proper	0.0889	0.135

3.2.4 Discussion

Supporting one hypothesis of this study, we find that DLMRI scores outperformed other MRI-based measures in both classifying prodromal AD and in predicting progression to dementia. While strongly influenced by changes in the entorhinal cortex and hippocampus, the DLMRI scores are distinguished from other MRI-based measures by incorporating the entire brain in its classification, agnostic of tissue type, volume or other pre-processing steps, likely accounting for its superior performance. Additionally, compared to measures of volume or thickness, the DLMRI scores are based on signal variability within individual voxels, and so is predicted to be more sensitive to synaptic loss, an earlier step in the neurodegenerative process (Khan et al., 2014).

Refuting another hypothesis, we found that DLMRI scores performed at least as well and typically outperformed biomarkers of amyloid and tau pathology in classifying prodromal AD. We do not consider this unexpected finding a challenge to the primacy of amyloid and tau pathology in the pathophysiology of AD (Clifford R Jack et al., 2013). DLMRI scores are found strongly linked to tau pathology in the entorhinal cortex, a region where AD pathology begins (Braak & Braak, 1991). The DLMRI scores' superior performance likely reflects this sensitivity. It is possible, therefore, that tau-PET would outperform DLMRI score and other biomarkers. ADNI has only begun acquiring tau-PET in 2015, and there is currently insufficient data to test this prediction in our experimental design.

The fact that DLMRI scores outperformed biomarkers of amyloid and tau pathology in predicting time to dementia prediction is less surprising. As a biomarker of neurodegeneration, this finding agrees with prior studies (S.M. Landau et al., 2010; Vemuri et al., 2009) and with the current model for the temporal sequence of AD pathophysiology (Clifford R Jack et al., 2013). Since neurodegeneration occurs last, accurate biomarkers of it are most proximal to the development of dementia.

The strength of this study is also its potential weakness. Relying on progression to AD dementia as a way to retroactively identify patients with prodromal AD sidesteps the limitation that precise biomarker cutoffs for prodromal AD are not yet known. We designed the analysis based on prior studies, that suggest that the majority of MCI patients with prodromal AD will progress within 4-5 years (Visser et al., 2006). Our analysis confirmed this assumption Figure A-4. Furthermore, approximately half of the MCI cohort ended up having prodromal AD, which agrees with previous approximations (Vos et al., 2015). Still, there are likely a minority patients in the stable MCI category who are harboring prodromal AD at baseline. The number of misclassified patients is likely to be low (Vos et al., 2015), and so this potential imprecision would not be expected to significantly alter our results. Tracking stable MCI patients for longer periods might address this concern, but would in fact raise a new one. Because of AD's high incidence in older populations, when tracking patients for a decade or more, some are expected to develop AD de novo after the baseline evaluation. At the very least, we can conclude that the results are beyond reproach for a 5-year time window after initial evaluation, a clinically meaningful epoch for both patients and health care providers.

Validating that deep learning can enhance the clinical utility of MRI is the most important implication of this study, we consider this study a proof-of-principle for deep learning, and not necessary for our specific version of it.

The current standard of care when evaluating a patient with MCI suspected of having AD is to order "neuroimaging studies" (Albert et al., 2011), most typically the conventional MRIs from which the DLMRI scores are derived. The rationale for this recommendation and its routine clinical implementation is not to "rule in" AD, but rather to exclude other non-neurodegenerative causes of dementia, such as strokes, bleeds, and tumors. Deep learning algorithms that can extract useful information from MRI's already acquired, for prodromal AD detection, have the additional advantage of reducing patient burden and cost incurred by lumbar punctures, injection of radioactive ligands, or other additional testing.

3.3 Age estimation in normal aging

3.3.1 Introduction

Age estimation is a way to study normal aging from a discriminative perspective, summarizing the age-related patterns of individuals and can also reveal the underpinnings that characterize the aging process. Besides normal aging, age factors into most diseases. Patterns that occur through the normal aging process must be accounted for when studying the diseases-specific patterns. The deviance of the estimated age from the chronological ages have shown utility in certain diseases, especially those thought to emulate advanced age state.

Brain age can be specifically defined as the estimated age using brain characteristics based on normal aging population. It can be regarded as fitting a complicated normal reference equation of age, deviance of which reflects abnormality. Brain age derived this way has been linked to education and self-reported physical activity (Steffener et al., 2016), and has been utilized in disease characterization including AD (Gaser et al., 2013), schizophrenia (Hugo G. Schnack et al., 2016), traumatic brain injury (Cole, Leech, & Sharp, 2015), etc., where deviation from normal aging trajectory implicates disease state.

There are three factors embedded in the definition of brain age: 1) the brain characteristics, 2) normal aging population, 3) estimation.

The brain characteristics can be derived from many sources, neuroimaging is the most common and comprehensive way to characterize the brain state *in vivo*. Within neuroimaging, there are studies using EEG (Electroencephalogram) (Al Zoubi et al., 2018), DTI (diffusion tensor imaging) (Mwangi, Hasan, & Soares, 2013), resting state BOLD fMRI (H. Li, Satterthwaite, & Fan, 2018), which focus on different aspects of brain states.

But the most common modality in brain age estimation is T1w structure MRI which reveal anatomical characteristics of the brain. Neuroanatomical characteristics are sensitive and robust biomarkers for aging process (Walhovd et al., 2011). Practically, as one most widely available and standardized neuroimaging modality, T1w structural MRI can be easily acquired for a large population. And within structural MRI modality, studies have utilized derived summary variables such as regional volumes or thickness (Valizadeh, Hänggi, Mérillat, & Jäncke, 2017), and also raw MRI scans (Cole et al., 2017).

And secondly, a dataset that best represent the studying sample is crucial since the model reflects the knowledge distilled from the training population. In this study, we propose using a dataset aggregated from publicly available multi-center neuroimaging datasets, which represent a diverse healthy population. This healthy study population is both in a larger scale than any of the age estimation studies before and is also carefully designed to be best suited for age estimation task with even age distribution across adult life span.

Lastly, with adequate study population and brain characteristics, the age is estimated based on statistical machine learning frameworks. Another way to formulate this problem is to extract generalizable features from the brain that can best capture the chronological age of a person given that the person is experiencing typical aging process that is present in general healthy population. Numerous traditional machine learning methods have been proposed for age estimation including relevance vector machine (Ashburner, 2007; Franke, Ziegler, Klöppel, & Gaser, 2010), Gaussian process (Gutierrez Becker, Klein, & Wachinger, 2018; Lancaster, Lorenz, Leech, & Cole, 2018), random forest (Konukoglu, Glocker, Zikic, & Criminisi, 2013), hidden Markov model (B. Wang & Pham, 2011), non-negative matrix factorization (Varikuti et al., 2018), etc.. Deep learning based methods are very suited for this task provided enough training data, and has been previously applied in (Cole et al., 2017) and demonstrated good performance.

Similar to the deep learning AD classification study, we also plan to explore the predictive regions in the regression task with ablation experiments and activation map based *post hoc* interpretation method.

In summary, we utilize a 3D deep convolutional neural network based regression model to estimate age using T1w structural MRI volumes from a diverse multi-study population that is sampled with even age distribution across adult age span. We demonstrate superior performance both in the hold-out test set from the same custom population and also in an independent life-span test set reported in previous studies.

3.3.2 Methods

We describe the population and the experimental setup used in this study. The MRI preprocessing steps, and the CNN (Figure 3-19) used to estimate age are very similar to the AD study, and are thus referring to sections 3.1.2.2 and 3.1.2.4. We also explored different variants of 3D ResNet models (He, Zhang, Ren, & Sun, 2015) ('resnet18', 'resnet34', 'resnet50', 'resnet101', 'resnet152') but did not show better performance. We then propose regionality analysis to interpret the model using ablation experiments and extending class activation map method.

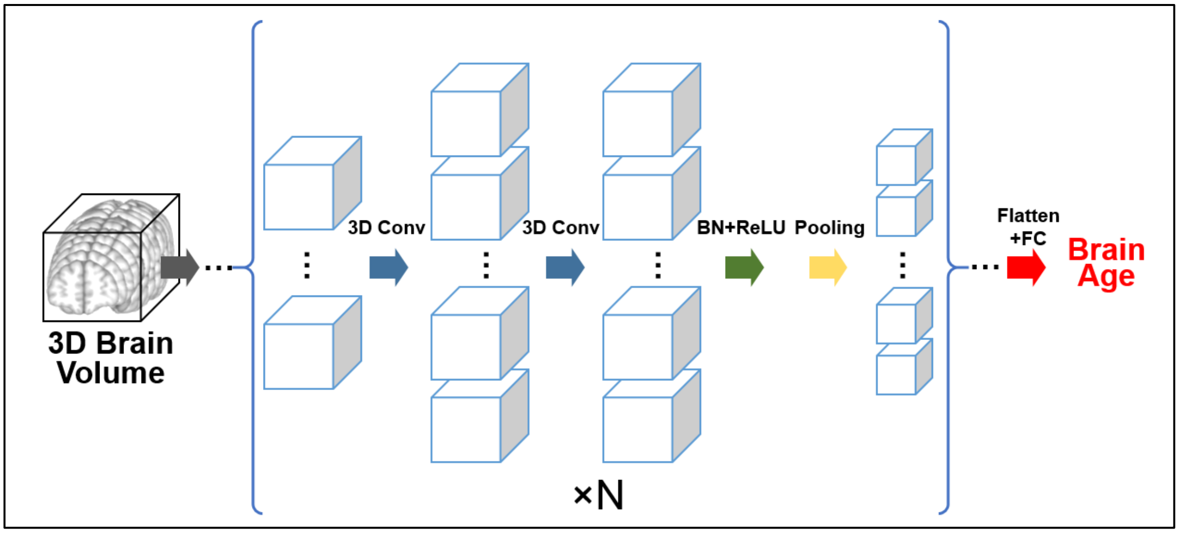


Figure 3-19. The convolution neural network architecture for age estimation. The only difference compared with Figure 3-2 is that the final convolutional output is flattened and fed into one fully-connected (FC) layer with *linear* output (red arrow).

3.3.2.1 Study Population

It is necessary to build an adequate neuroimaging dataset for age prediction in the full adult life span, especially given study recruitment criteria that are not necessarily evenly distributed. Recent progress in open imaging dataset greatly facilitates the possibility. In this work, we collect more than 30,000 T1w MRI scans from multiple open neuroimaging datasets. The list of the datasets used in this study with the full names and sources are listed in Table A-1. Among those, we only include subjects with clear indication of normal neurological evaluations. Basically, we exclude subjects with any neurological or psychiatric disease, and also subjects with no clear diagnosis label documented. We also chose 18 as the minimum age to cover the adulthood and also to avoid the neurodevelopment stage.

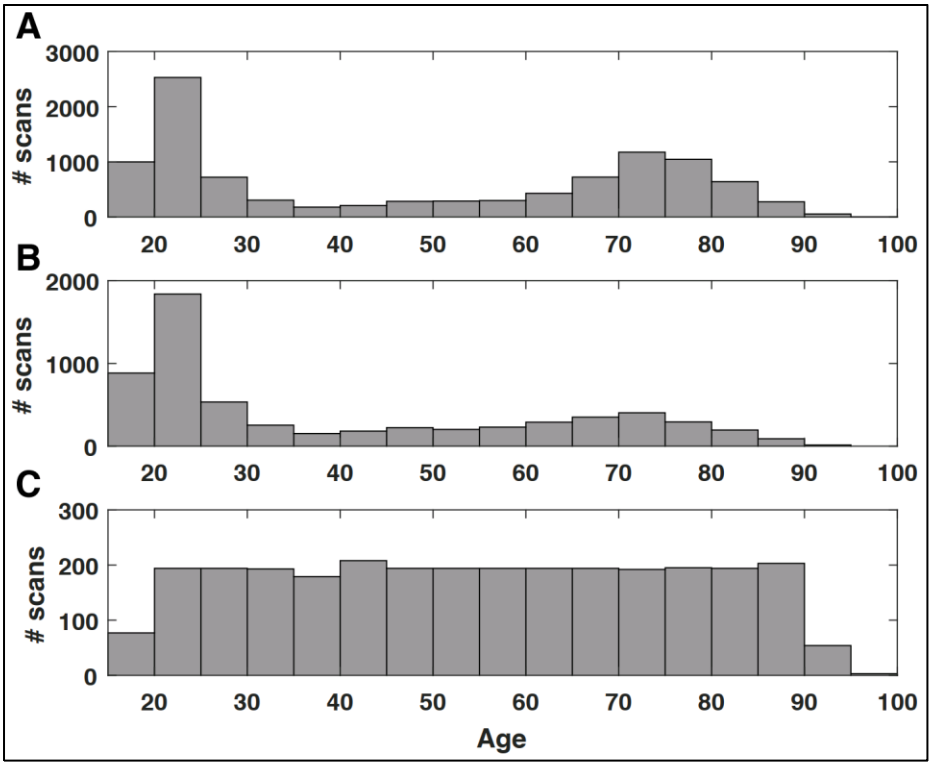


Figure 3-20. The age distribution of the age estimation study population. A) The age distribution of the raw dataset consisted of 10,158 scans; B) the age distribution of the dataset consisted of 6,142 unique subjects; C) the age distribution of the evenly sampled dataset.

This results in 10,158 MRI sessions, reflecting one of the largest and most heterogeneous structural brain MRI study ever conducted, the statistics are summarized in Table 3-4. However, as shown in Figure 3-20 (A), where we illustrate the raw age distribution, the age distribution of the population is highly unbalanced. Although there are studies covering the full age span including normal aging studies Cam-CAN (Taylor et al., 2017), IXI, SALD (Wei et al., 2018), DLBS (Rodrigue et al., 2012), OASIS-1 (Marcus et al., 2007), and consortium based studies such as CoRR (X.-N. Zuo et al., 2014), SchizConnect (L. Wang et al., 2016), many of the public imaging studies either focused on age-related disease in the elder population including ADNI, AIBL (Ellis et al., 2009), OASIS-2 (Marcus, Fotenos, Csernansky, Morris, & Buckner, 2010), PPMI (Marek et al., 2011), NIFD; or focused on young subjects including BGSP (Holmes et al., 2015), SLIM (Liu et al., 2017).

Table 3-4. Multi-center datasets used in deep learning and age estimation.

Dataset	$\mathbf{N_{sessions}}$	$\mathbf{N_{subjects}}$	**age range**	$\mathbf{age_{session}}$ $\mathbf{mean_{\pm std}}$	$\mathbf{gender_{session}}$ **M/F**	$\mathbf{gender_{subject}}$ **M/F**
ADNI	2423	438	56.3 - 95.8	$76.64_{\pm 6.05}$	1223/1192	217/221
AIBL	781	457	60 - 92	$72.86_{\pm 6.53}$	359/421	199/258
NIFD	428	136	36.9 - 85.2	$65.76_{\pm 7.66}$	185/243	59/77
IXI	561	561	20.0 - 86.3	$48.67_{\pm 16.47}$	248/313	248/313
BGSP	1566	1566	19 - 35	$21.54_{\pm 2.89}$	661/905	661/905
Cam-CAN	652	652	18 - 88	$54.30_{\pm 18.59}$	322/330	322/330
OASIS-1	316	316	18 - 94	$45.09_{\pm 23.90}$	119/197	119/197
OASIS-2	145	56	60 - 97	$76.54_{\pm 7.99}$	52/93	19/37
SALD	467	467	19 - 80	$45.07_{\pm 17.39}$	168/297	168/297
SLIM	972	561	18 - 28.5	$20.68_{\pm 1.40}$	425/547	244/317
PPMI	130	74	30.6 - 81	$60.95_{\pm 10.85}$	91/39	48/26
SchizConnect	742	567	18 - 70	$35.36_{\pm 12.76}$	429/313	338/229
DLBS	301	301	20.6 - 89.1	$53.62_{\pm 19.92}$	115/186	115/186
CoRR	1326	642	18 - 83	$28.78_{\pm 12.39}$	689/637	343/299
total	10158	6142	18 - 97	$47.92_{\pm 24.89}$	4764/5783	2780/3362

To alleviate the potential bias toward a certain age group, we need to balance the age distribution in the training population. In this study, we adopt both oversampling and undersampling. Specifically, we oversample age ranges with less subjects by including the longitudinal follow-up sessions from the same subjects, which could be regarded as a natural augmentation. For age ranges with more subjects, we only include one scan per subject to increase the variability of the sample, and if the number is still much larger, we further undersample stratified on study and gender.

We divided the populations into age bins and use the bin with the minimum number of subjects as the basis number. The age bins used in this study are [18, 20), [20, 25), [25, 30), ... , [85, 90), [90, 100). One interesting observation is that the [35, 40), [40, 45) are the two age ranges with fewest number of subjects, so we regard these two bins as one, and allowing repeated scans from same subjects, and use the number of scans as the base level. The other age bins having multiple scans per subject are [85, 90) and [90, 100] because of the relative lower number in the two age bins. We undersample the subjects in other bins. The final dataset consists of 2856 MRI sessions from 2694 subjects covering age range 18-97, with the mean age 54.34 years old, standard deviation 21.16 years old. The age distribution of the evenly sampled adult age span dataset is shown in Figure 3-20 (C).

3.3.2.2 Experimental setup

We design training, validation and test sets of subjects for model training and evaluation. Since validation set is used to choose the optimal parameters of the model, we need to build a diverse validation set reflecting general MRI scan distribution and population distribution. Similarly, a test set representative of the same population as the training and validation sets is important for evaluation. Given the even age-distributed dataset we described in the section above, we perform stratified split based on study and gender within each age bin: 8/10 as training set, 1/10 as validation set, 1/10 as test set, ensuring non-overlapping subjects and similar distribution of age, study, gender in the training, validation and test sets.

We also evaluated our model by testing it on an independent test set. In this study, we use the Cam-CAN study as an independent test sample. Aiming to study normal aging process, Cam-CAN provided very even age-distribution across adult life-span and has been previously used as an independent testing sample in (Lancaster et al., 2018). We have to point out that, ideally, the trained model should be able to apply to all kinds of population. But the result on an independent sample which usually reflects a homogeneous population or acquisition setting, might be over- or under-optimistic.

We also perform a test-retest experiment using an independent dataset of three subjects scanned 40 times in 30 days (Maclaren, Han, Vos, Fischbein, & Bammer, 2014) to test the reproducibility of the model.

3.3.2.3 Comparison with model trained on unbalanced dataset

We also trained the model using all scans from unique subjects (N = 6,142, Figure 3-20 B). We apply the trained model on the independent Cam-CAN dataset and study the distribution of MAE over chronological age groups.

Besides, a simple way to potentially correct the imbalance without adjusting the sampling of the dataset is re-weighting the samples, specifically, we assign different weights to different sample, with the weights in proportional to the inverse of the frequency of specific age segments.

3.3.2.4 Neuropsychological and morphometric associations

To test the utility of the estimated age in studying cognitive functions across adult life-span, we evaluate the association between the summary scores of Benton face recognition test (BFRT) and the estimated age in Cam-CAN dataset.

Specifically, we use the difference of the estimated age and chronological age to measure the deviance of individual brain age from their chronological age, and we refer to this value as age_{diff} hereinafter.

The BFRT is a commonly used neuropsychological instrument that can be easily and reliably administered in adult patients to test baseline visual memory and perception. We adopted the SubScore-1, SubScore-2, TotalScore (SubScore-1 + SubScore-2) as dependent variables in individual linear regression models incorporating gender, chronological age, age_{diff}, and the interaction of chronological age and age_{diff}:

$$\text{Score} \sim \beta_{\text{age}}\text{age} + \beta_{\text{age}_{\text{diff}}}\text{age}_{\text{diff}} + \beta_{\text{age}_{\text{diff}}\times\text{age}}\text{age} \times \text{age}_{\text{diff}} + \beta_{\text{gender}}\text{gender}$$

Additionally, we evaluate the association between age_{diff} with cortical thickness generated using FreeSurfer (Bruce Fischl & Dale, 2000), by performing a partial correlation with gender and chronological age as covariates. FreeSurfer parcellates the cortex into 68 regions, and can be combined bilaterally into 34 regions.

3.3.2.5 Age activation map

Class activation mapping (Selvaraju et al., 2017; Zhou et al., 2016) reflects a commonly used method for interpreting the classification using CNNs and has been previously used in CNN based medical image analysis (Feng et al., 2017) to marry potential disease pathology with classification findings. In this work, we use the idea of a class activation map in a regression setting by highlighting the small-valued gradient in grad-CAM framework. We used functions in the keras-vis package[11]. We generate the average activation map within each age group to investigate the age-specific pattern of underlying substrates for age estimation.

3.3.2.6 Slice based age estimation

Besides the *post hoc* saliency map based activation map method, we also propose ablation analyses methods focusing on part of the input data. We apply serial 2D CNNs for age estimation with the input being three consecutive slices along each axis. The network architecture of the 2D CNN is the identical to the 3D CNN architecture described in the previous section with the 3D operations replaced with the corresponding 2D operations. We report the estimation performance on the different set of 2D slices as the indication of predictive importance.

3.3.2.7 Lobe based age estimation

Besides sliced based age estimation, we proposed using another more neuroanatomically-informed way to study the regionality through ablation experiments at the lobe level. The individual lobe masks were generated following the section 3.1.2.8 or our previous study (Feng, Yang, et al., 2018). The ages were estimated focusing on each lobe individually.

---

[11] https://github.com/raghakot/keras-vis/

3.3.3 Results

3.3.3.1 Age prediction

In the hold-out test set from the same population distribution as the training and validation set, our model achieves an MAE of 4.06 years and r = 0.970. For the independent Cam-CAN test set, the model proposed in (Lancaster et al., 2018) achieved MAE = 6.08 years and r = 0.929 when the Cam-CAN data were pre-processed with the optimal parameters selected from the independent training sample. We tested out our model in Cam-CAN data, processing the T1w MRI images using the proposed pipeline. The relationship between the estimated age and chronological age in Cam-CAN is shown in Figure 3-21, where r = 0.96, MAE = 4.21 years, which outperforms the result in (Lancaster et al., 2018). We also observe two obvious "outliers" among the 652 subjects tested, further investigation is needed to pinpoint whether they are due to methodological reasons or pathological reasons. The results demonstrate our proposed model achieve an accurate estimation in all age segments.

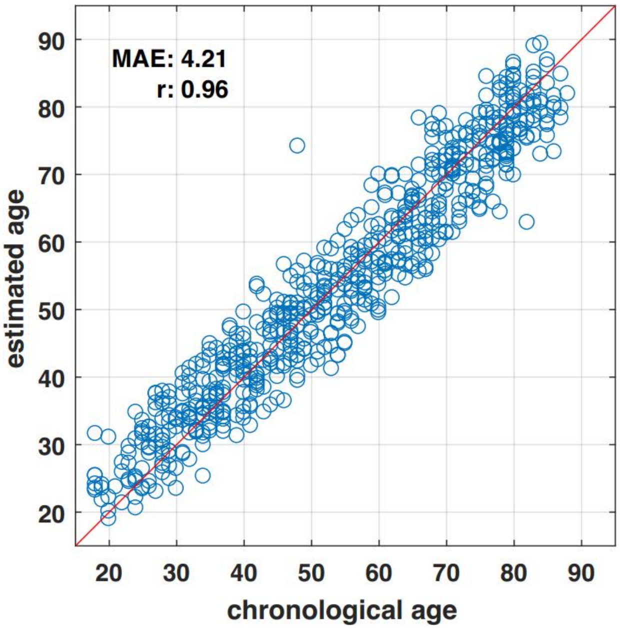


Figure 3-21. The estimated age vs. chronological age in an independent test set.

3.3.3.2 Reproducibility

We evaluate the reproducibility of the algorithm in test-retest scans. We show the statistics in Table 3-5 and Figure 3-22. We observe that there is a difference in the estimated age and actual reported age. But the difference is consistent over the sessions with around 1 year old standard deviation, demonstrating good reproducibility of the proposed model.

Table 3-5. Test-retest reproducibility experiment results.

Subject	Actual age	Estimated age mean	Estimated age std
Subj-1	26	25.19	1.07
Subj-2	31	33.02	1.14
Subj-3	30	27.06	0.81

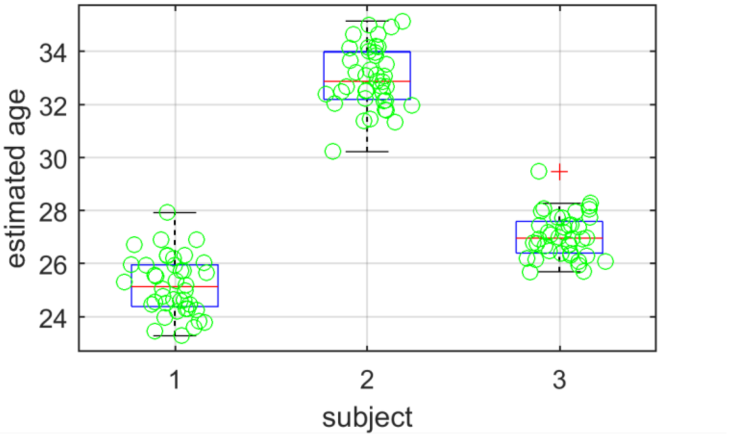


Figure 3-22. The distribution of predicted ages in test-retest scans.

3.3.3.3 Comparison with results using unbalanced dataset

We compare the results using unbalanced dataset with the MAE performance in Cam-CAN dataset. Using the unbalanced dataset achieves comparable overall MAE (4.27 years) as the balanced data. Re-weighting the samples helps slightly improves the MAE (4.17 years) than the balanced dataset despite using many more scans. Additionally, we observe the MAE using the unbalanced dataset is not evenly distributed across life-span: MAE is lower in the young age with more abundant data, as shown in Figure 3-23 (B).

This could introduce potential bias in life-span studies. Using sample re-weighting (Figure 3-23 (C)) alleviates the problem, and using balanced dataset generates even distribution across age-span (Figure 3-23 (A)).

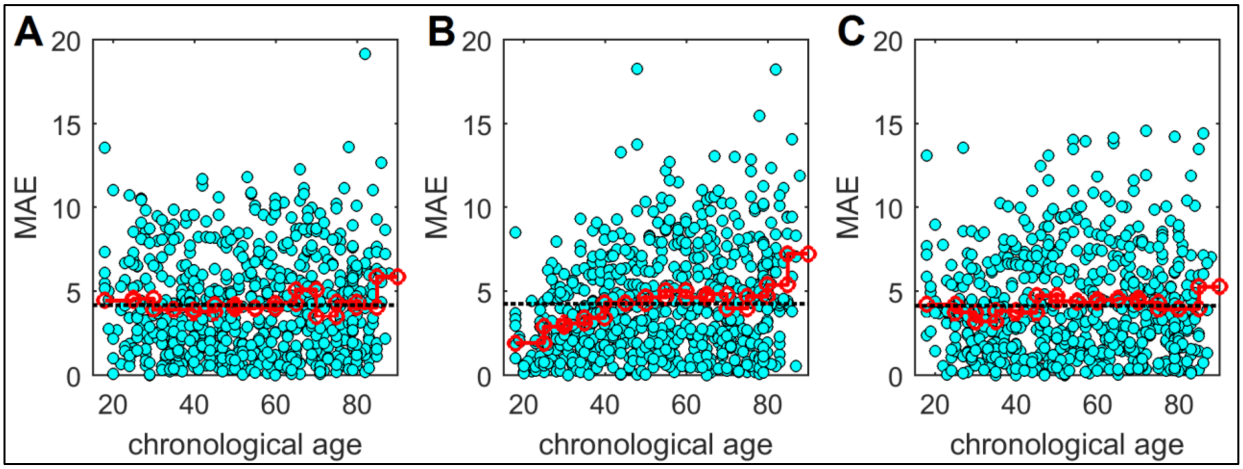


Figure 3-23. Distribution of MAE of age estimation across life-span. (A) Age estimated using the balanced dataset. Each step in the red line indicate the MAE in that age group, the black dashed line indicates the overall MAE. (B) Age estimated using the unbalanced dataset. (C) Age estimated using the unbalanced dataset but with sample re-weighting.

3.3.3.4 Neuropsychological and neuromorphometric association

The association of the cortical thickness measures with the age_{diff} is illustrated in Figure 3-24. The thickness of cortical regions are significantly associated with the age_{diff}. In addition, out of the 68 regions measured, 51 regions have a stronger correlation with the estimated age than the chronological age. This is expected as the age estimated through structural MRI image is in principle more coupled to structural phenotypes.

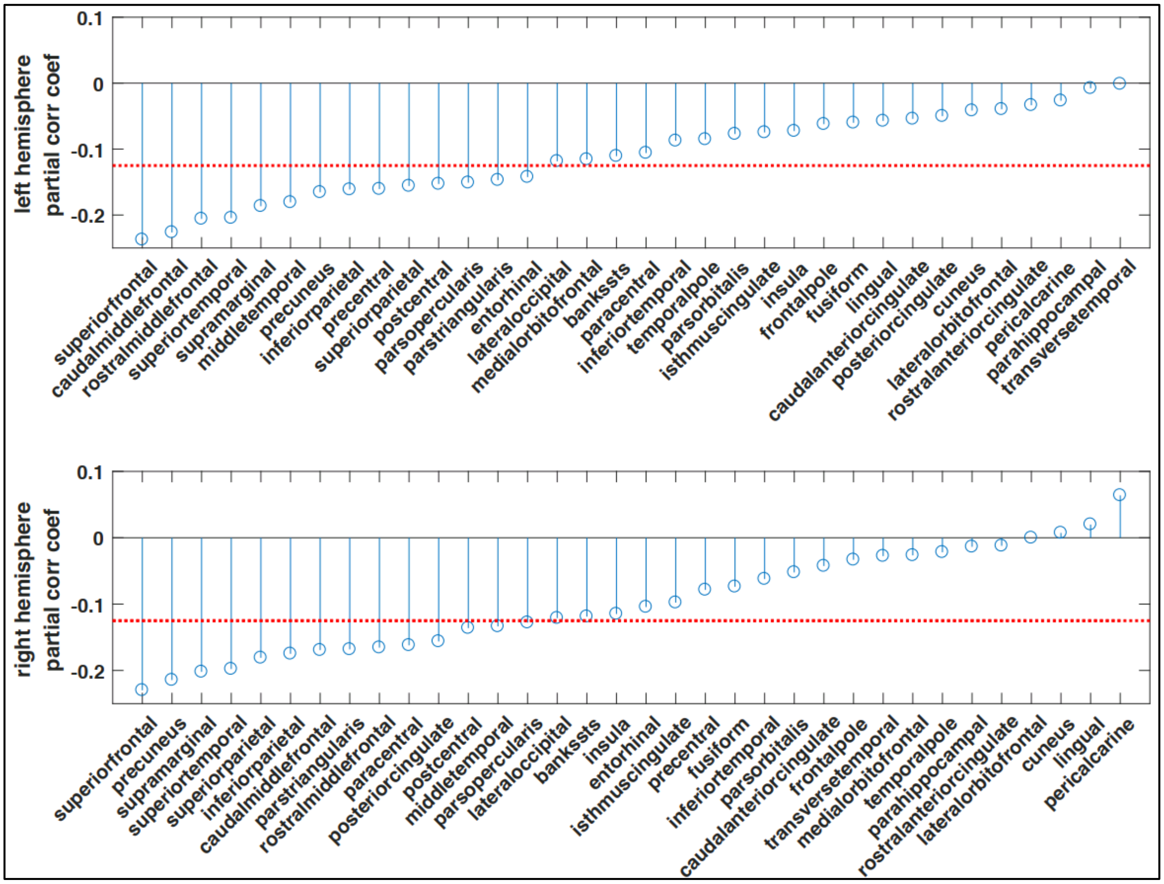


Figure 3-24. The partial correlation coefficients of age_{diff} and cortical thickness measures. The red dashed line indicates α under multiple comparison of N=34 regions.

The association of the BFRT scores with the difference in the estimated age and the chronological age, and its interaction with chronological age are summarized in Table 3-6.

Table 3-6. Association with Benton face recognition scores

Score	age		age diff		age diff × age	
	β	p	β	p	β	p
SubScore-1	-2.51e-3	3.01e-6	-0.0127	0.0297	2.62e-4	0.0102
SubScore-2	-0.0605	1.24e-32	-0.157	2.90e-3	2.31e-3	0.0119
TotalScore	-0.0630	5.07e-34	-0.170	1.54e-3	2.57e-3	5.88e-3

3.3.3.5 Age activation maps

It is expected the anatomical patterns characterizing different age groups to be different throughout lifetime but are consistent within a local age range. Thus the age activation maps were generated and illustrated every 5 years, in the same way as preparing the dataset. The 3D iso-surfaces of the average age activation maps are shown in Figure 3-26. The average age activation maps overlaid on the MNI152 template are shown in Figure 3-25 (Left). To accommodate the anatomical differences in different age groups, average T1w images within each age group were generated, and the corresponding age activation maps were overlaid, as shown in Figure 3-25 (Right). In the 20s, the predictive region localizes to orbital-frontal regions, which are the regions having most significant age-related volume decline across life-span as shown in Figure 2-17 in the previous aging chapter. With increasing age, the predictive regions still localize to the vicinity, but there is a trend moving superiorly in 30s - 60s. In 60s - 80s, the pattern stays at similar location but becomes more diffuse (revealed as smaller iso-surface). In the 80s, the overlapped areas are mostly anterior cingulate cortex and lateral ventricle.

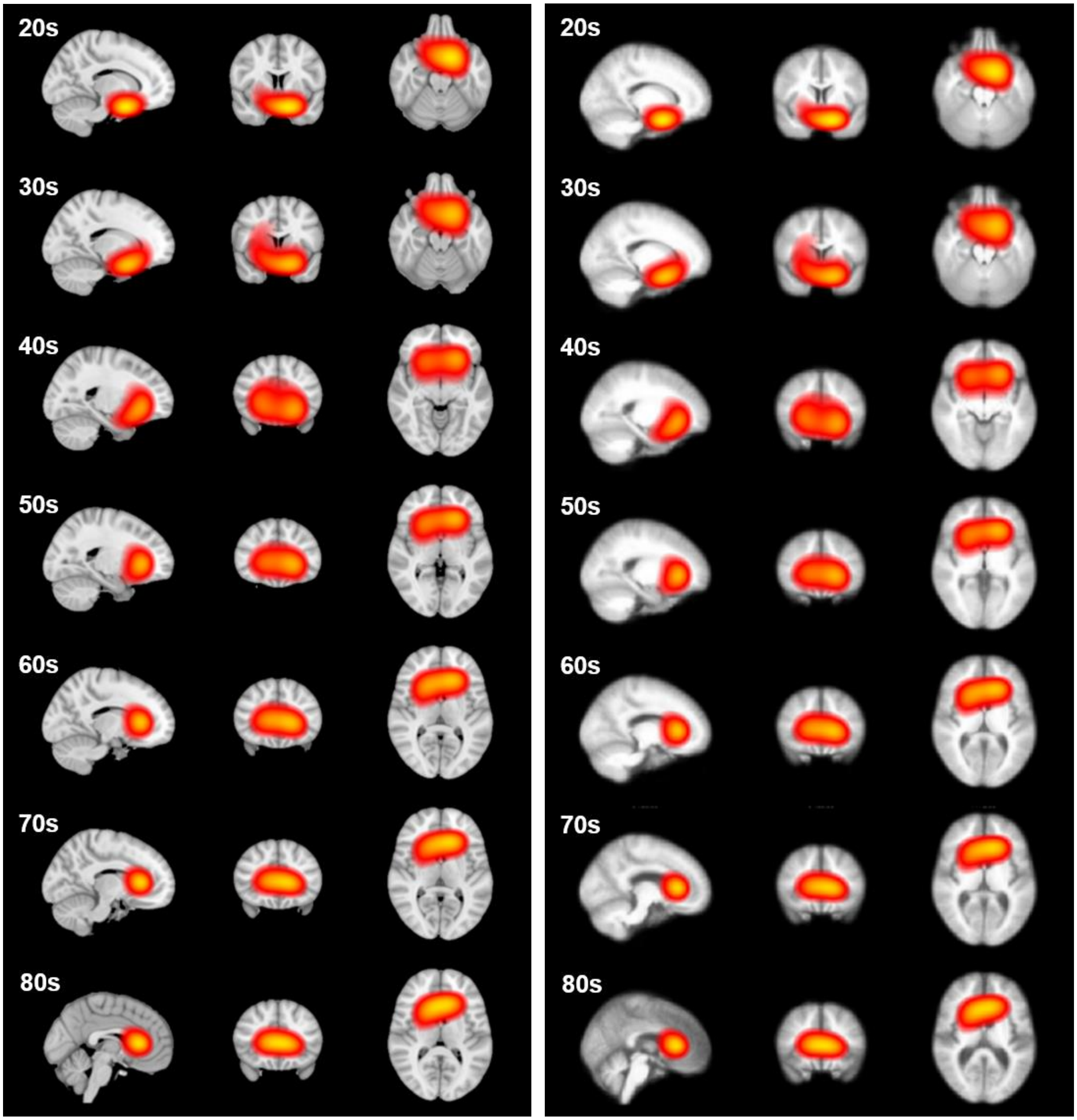


Figure 3-25. The age activation maps at different age groups. The age activation maps overlaid on the (Left) MNI152 template, and (Right) average T1w image within each age group, both with threshold at 0.8.

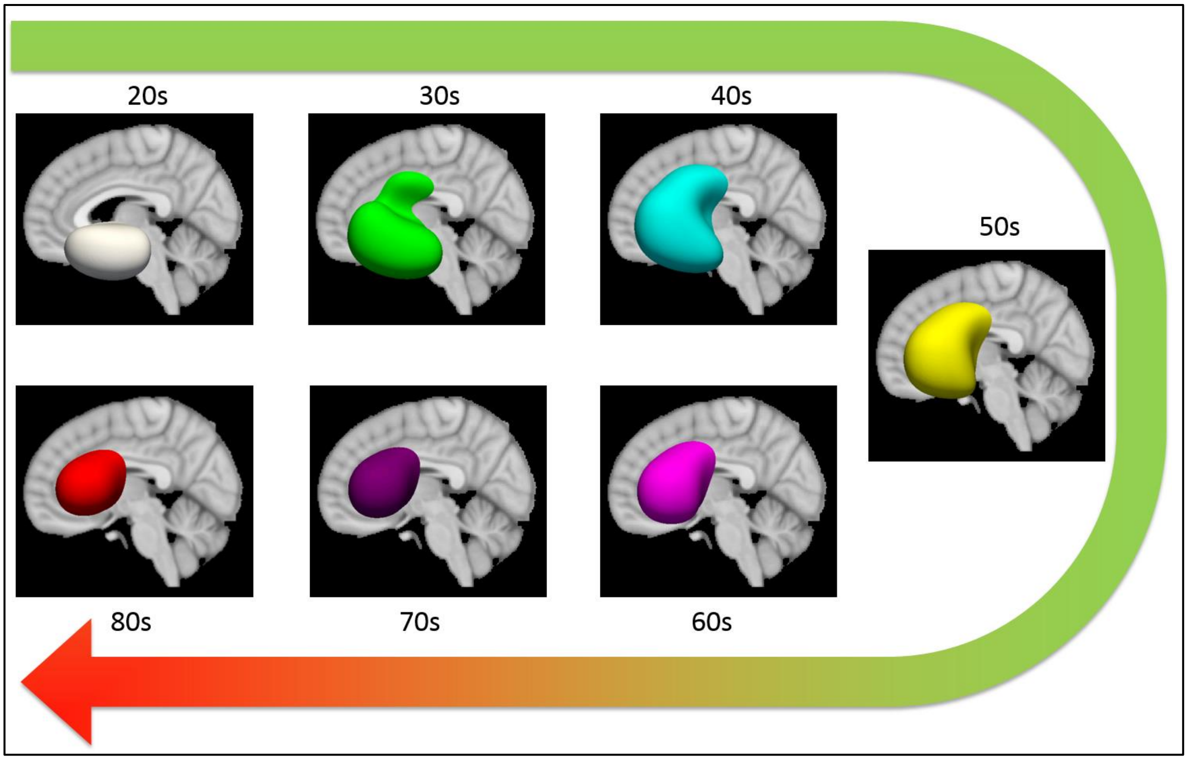


Figure 3-26. The 3D iso-surfaces (0.8) of the age activation maps at different age groups.

3.3.3.6 Slice based age estimation

The age estimation performance using 2D MRI slabs sliced at different coordinate planes is shown in Figure 3-27. The slices with the best performance are also illustrated. The patterns are more diffuse. But from the results in the coronal orientation, the anterior side or the frontal regions are more predictive.

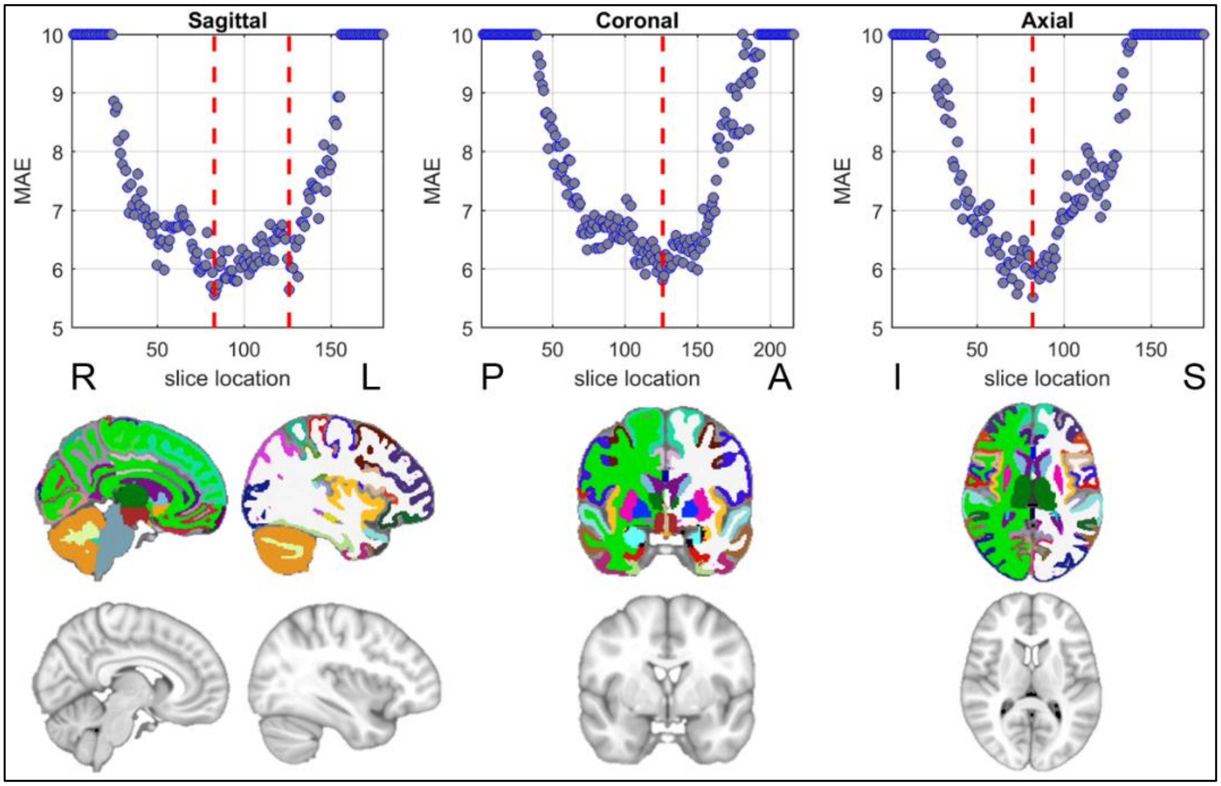


Figure 3-27. MRI 2D slice based age estimation. (Top row) The mean absolute error (MAE) of the estimated age on the test set using 2D slices at different locations, the red lines indicate the location with lowest MAE. MAEs larger than 10 are cut to 10 for illustration purpose. (Bottom row) The illustration of slices at the red line in the top row from the MNI152 template and the corresponding segmentation (the colors follow the FreeSurfer color lookup table).

3.3.3.7 Brain lobe based age estimation

The age estimation performance of models trained on different lobes and cerebellum are shown in Table 3-7, the model trained on the frontal lobe achieves the best age estimation performance, but inferior to the performance using whole brain. And temporal lobe achieved marginally lower performance.

Table 3-7. Age estimation performance achieved using individual lobes and cerebellum.

Lobe	Frontal	Temporal	Parietal	Occipital	Cerebellum	Whole-brain
Age MAE	5.33	5.81	6.37	7.66	6.20	4.06

3.3.4 Discussion

In this study, a large heterogeneous dataset of structural neuroimaging across the adult lifespan was aggregated from multiple publicly available data sources, from which we curated a uniformly distributed dataset, which is able to achieve estimation not biased toward certain age group, while maintaining training efficiency comparing with training with all the data weighted according to the abundance.

3.3.4.1 Regionality

The regionality analysis in this study revealed patterns of neuroanatomical contributions of normal aging. All analyses provide evidence for the prominence of frontal regions in age estimation. Frontal regions have been shown to be selectively implicated in normal aging through both neuropsychological studies and neuroimaging studies (Chetelat et al., 2013; Gazzaley et al., 2005; Shamchi et al., 2017). In addition, the pattern shifts reflected in the class activation map based analysis imply the complexity of age-related structural changes. Neuropsychological evaluations targeting different cognitive domains and brain regions could help reveal the complexity in the aging-process, and might be required to evaluate different aging stages.

3.3.4.2 Open questions

One limitation comes from the definition of healthy subjects. The set of subjects might include subjects with pre-clinical AD or other disease at prodromal state that were not manifested in the clinical evaluation. However, this is not easily resolvable for cross-sectional studies.

Our analysis revealed the association between the divergence of estimated age from chronological age and BFRT performance. This suggests the potential utility of the estimated age at normal aging evaluation, in complement to other cognitive test and neuroimaging based measures. The utility relates to an open question of the aging process. It still requires further validation how the deviance would tell an individual's brain health status or even trigger clinical evaluations, since inter-subject differences in normal aging process definitely exist but are not well understood. And it is unclear whether this inherent variability increases with age or stays constant.

3.3.4.3 Study sample composition

Finally, while this study aims to study adult life-span, other studies aiming at different neurological conditions might benefit from other training dataset inclusion criteria, for example, autism and prodromal psychosis studies would definitely require the inclusion of subjects below 18, and subjects in middle and old age are less informative for the condition being explored.

3.3.4.4 An overarching framework

The Alzheimer’s disease diagnosis and age estimation studies constitute the key parts of a deep learning powered neurological condition evaluation workflow using T1w structural MRI. The workflow is designed with the inspiration from general neurological practice. In the deployment scenario, given a new MRI from scanner, quality check and basic preprocessing will first be performed. And then the preprocessed image is fed into different models. At this stage, the deep learning inference can be done in less than a second for each model. With more data and more models, we can derive more variables. Taking together the wide availability of T1 weighted structural MRI, this pipeline has special practical advantage, and could be especially useful given limited resources.

One of the implications of supervised learning is that the desired properties of the model are largely determined by the input data and the associated labels. Thus, the workflow parallels the introduction of available datasets, and proposes potential future plans of data collection to complete the system. The specific algorithm also depends on the deployment scenario whether might it be inclusion/opt-in or exclusion/opt-out.

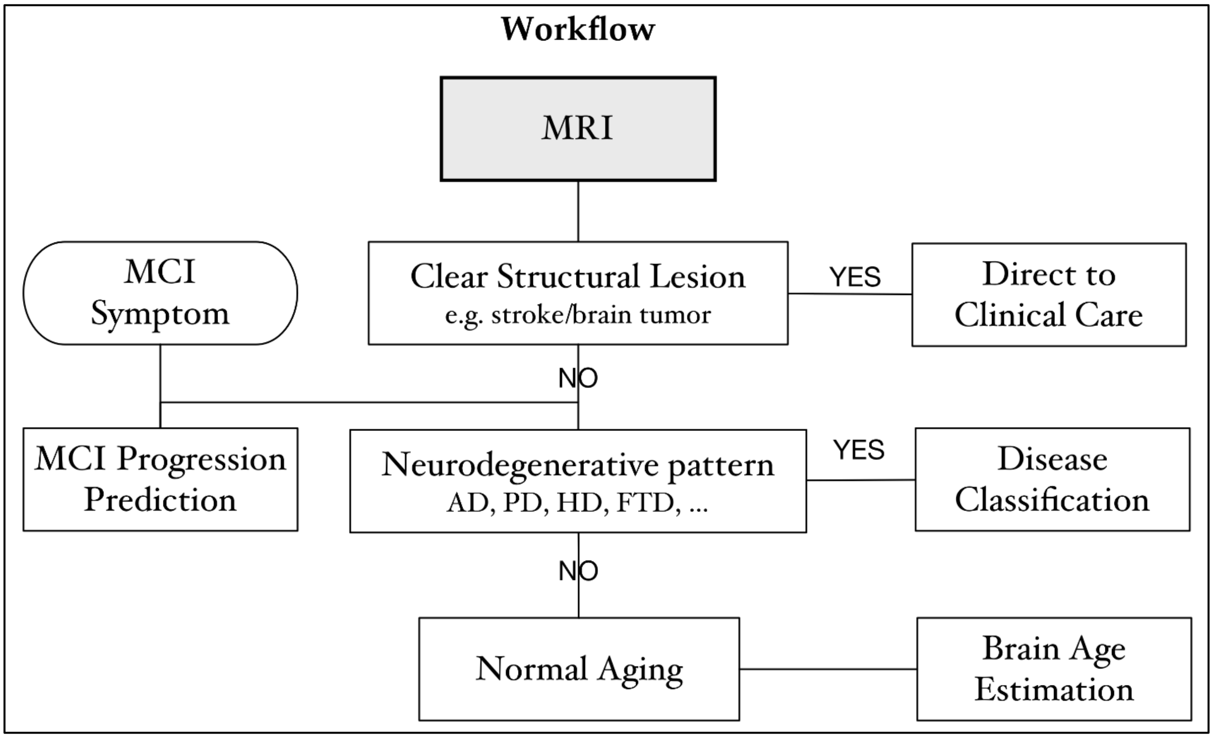


Figure 3-28. A framework of workflow in neurological condition evaluations.

The first step is to detect clear structural lesions such as brain tumor and stroke. This can be a supervised learning task using available datasets on brain tumor such as BRATS (Brain Tumor Segmentation) (Menze et al., 2015) from The Cancer Genome Atlas (TCGA) collection and stroke such as ATLAS (Anatomical Tracings of Lesions After Stroke) (Liew et al., 2018). These datasets also include manual segmentation of the lesions from experts. Thus, another way of utilizing these "dense" annotations is to train a stroke/brain tumor segmentation model. If no obvious (large cluster of) segmentation is predicted when applied on the new incoming data, the brain deems patent-lesion free. Another potential solution can be based on an unsupervised anomaly detection method, but is beyond the scope of current study. Patients with clear structural brain lesion would then be directed to clinical cares for further evaluation.

The next step is to test the existence of anatomical pattern of neurodegenerative disease. The ideal training dataset should consist of patients with different kinds of neurodegenerative disease versus healthy controls. If the anatomical pattern falls into neurodegenerative disease pattern, the next step is to further classify the specific disease including Alzheimer's disease (AD), Parkinson's disease (PD), Huntington's disease (HD), frontotemporal dementia (FTD), Lewy Body Dementia (LBD), etc… And it has to be noted that some of the diseases can be further divided into subtypes, and certain diseases have overlapping pathological anatomical patterns, and might only be differentiated with the information from other data modality.

In section 3.1, I focused on Alzheimer's disease diagnosis. Although not specifically designed to be specific to Alzheimer's disease, the algorithm shows specificity in HD, PD, schizophrenia, and certain subtypes of FTD. In the future, it is possible to train a multi-class model with AD, control and other non-AD neurological disease. When a patient comes to the clinic and presents MCI symptoms, the model can classify prodromal AD and evaluate time to conversion to AD.

If a subject is free of clear structural lesion and neurodegenerative pattern, then the subject lies in the spectrum of normal aging structurally. The location within the normal aging spectrum can be evaluated using the age estimation model.

The same set of system can be easily generalized to other MRI contrasts such as CBV, CBF or other imaging modalities such as PET, SPECT, CT. The model can always be used as a pattern extraction/recognition system with the probability simply interpreted as the similarity to a specific pattern (e.g. AD pattern). So it can always be used together with other sources of evidence such as age, gender, functional imaging measures, neuropsychological measures, CSF biomarkers, etc.

4 Specific Aim 3: Enhancing the utility of CBV with retrospective CBV

In this section, I introduced a technique enabling CBV generation retrospectively from clinical contrast-enhanced scans. The derivation of meaningful functional measures from such clinical scans is only possible through calibration to a reference, which was built from the *largest* collection of research CBV-fMRI scans from our lab. The method was validated in an epilepsy study and has demonstrated the potential to enhance the utility of CBV-fMRI by enriching the CBV-fMRI dataset. This technique is also applicable to AD and normal aging studies, and potentially enables deep learning based analytic approaches on CBV-fMRI with a similar pipeline used in structural MRI.

4.1 Introduction

Contrast-enhanced-MRI (CE-MRI) is a valuable and well-established diagnostic imaging tool used clinically to reveal signal changes reflective of oncological or inflammatory origin, as well as other underlying pathologies. CE-MRI is used annually in ~30 million procedures, and more than 300 million procedures have been performed as of 2015 since 1988 (Lohrke et al., 2016). Many of the protocols for these scans overlap with the protocols of the requisite scans necessary for research CBV.

At its core, what distinguishes CBV-fMRI from other fMRI measures is the introduction of contrast agents, which provide higher spatial resolution and increase in intravascular signal to noise ratio (SNR) (Lin et al., 1999). Moreover, the majority of clinically ordered CE-MRI reflect intact vasculature (radiologically noted as “non-contrast enhancing”). As such, the signal changes after contrast enhancement reflect the passive (venule and capillary) component of blood capacity, same as the research CBV-fMRI.

However, in a clinical setting, dynamic ranges of the signal in pre-contrast and post-contrast acquisition may vary, especially if parameters related to the MRI voxel intensity values are left to scanner default parameters (as is common for MRI scanners unless otherwise specified), resulting in *scaling* problem. Whereas in research CBV-fMRI scans, the post-contrast and pre-contrast image are in the same intensity translation and scaling system to accurately calculate the contrast uptake and determine voxel level CBV values. In certain experimental practices, an external intensity fiducial marker may be scanned at the same time to help with intermodal co-registration (Woods, Mazziotta, & Cherry, 1993), where the intensity of the fiducial marker may be used to correct gross image intensity differences (Simor et al., 1995). But for most scan sessions, there are no external fiducial markers. In these scenarios, there is not a straightforward way to correct the image for potential quantitative CBV processing. But we can borrow the idea of external fiducial marker by finding an internal fiducial marker and validate the fiducial marker in a dataset with known dysfunction.

In this case study, I found an internal reference with the guidance from the largest collection of research CBV-fMRI, and explore this method on a unilateral temporal lobe epilepsy (TLE) lateralization task using MRI scans acquired for surgical evaluations. For TLE patients who are potential candidates for surgical treatment, lateralization of the epileptogenic zone is essential, and presents a challenge in the context of an unremarkable clinical structural MRI scan. Temporal lobe epilepsy laterality has been studied using many different imaging modalities including PET (Theodore, Sato, Kufta, Gaillard, & Kelley, 1997), ASL (Wolf et al., 2001), and DSC-MRI (Wu et al., 1999) for metabolism studies and structural MRI for structural studies (Mueller et al., 2009).

I interrogated the sub-regions of the hippocampus within this TLE group and potentially provide insight into the pathophysiology of the hippocampal circuit in epilepsy using an approach that can be retrospectively and prospectively applied to patients who would receive these scans regardless.

4.2 Methods

4.2.1 Data

We included 304 pairs of correctly scaled pre-contrast scans and post-contrast scans in the study acquired from a collection of studies, which cover a broad population of subjects across different studies and age ranges, including follow-up scans of the same subjects. Both healthy subjects and subjects with various diseases, including schizophrenia, post-traumatic stress disorder, Alzheimer's disease, are included. The scans were acquired using CBV-fMRI sequence as described previously in normal aging study. Those reference scans are used to determine an appropriate reference region.

The research scan database is used to identify which class or region could be used as an internal reference region, as shown in the top. That finding is then applied to our clinical research scans shown in the bottom. For the research stream, brain extractions, regional and tissue segmentation are performed on the pre-contrast T1 weighted image. Post-contrast images are co-registered to pre-contrast images and pre and post ratios are generated arithmetically on a voxel by voxel basis. For the clinical database, post-contrast scans are also co-registered in the same manner as the research database, in addition to brain extraction and tissue segmentation. White matter post/pre ratios, as further described in the text, were used to generate the scaling correction factor (SCF). The SCF was then used to correct the post-contrast image which was used to generate the properly scaled CBV scans.

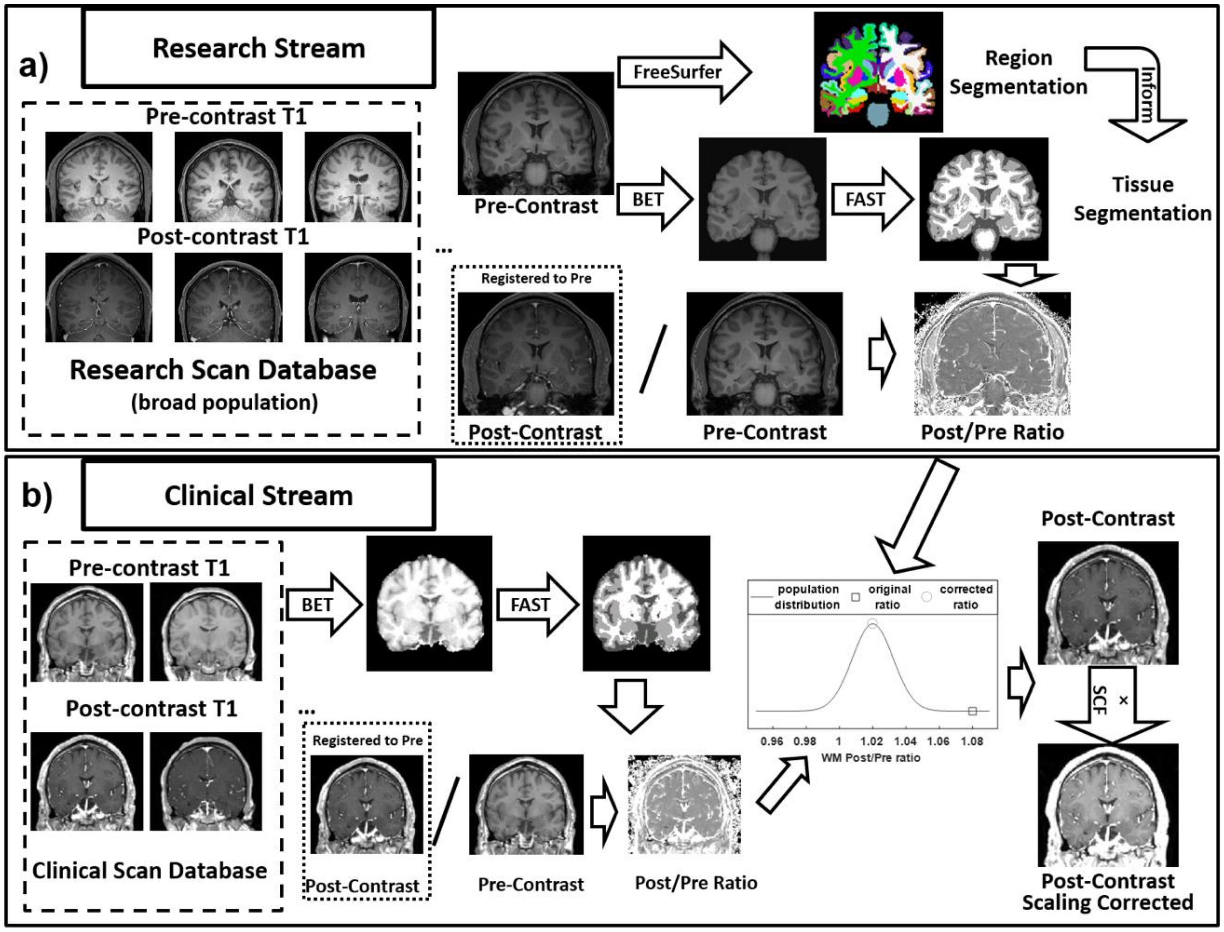


Figure 4-1. The retrospective CBV pipeline.

For the retrospective clinical study, 34 unilateral TLE patients with pre-contrast and post-contrast T1-weighted scans were screened from an internal database via retrospective chart review from 2011 to 2016. The laterality was determined based on video-scalp EEG recording of seizure onset with no evidence of seizures arising from the contralateral hemisphere. Other inclusion criteria include age>18 and IQ>70. Predicting a similar level of effect, the number of patients included exceeds an existing study demonstrating potential hypometabolism effect (Wolf et al., 2001). Four patients were excluded from study due to space-occupying hippocampal lesions as noted by a radiologist through clinical evaluation. Four additional patients were excluded due to MRI technical issues including significant mismatch of MRI acquisition parameters of pre and post-contrast scans (one case: pre-contrast image TE 3ms TR 16.67ms, post-contrast image TE 5ms TR 19ms), severe artifact, prohibitive field of view or inadequate segmentation. The exclusion process was blind to patient diagnosis information. The receiver gain calibration was on all scans. Mesial temporal sclerosis (MTS) is not an exclusion criterion in this study, 5/26 patients have MTS.

4.2.2 Scaling correction factor estimation with brain region segmentation

A diagram demonstrating both the research and clinical databases and streams is shown in Figure 4-1. The first step in attempting to identify a way to correct scaling within clinical scans is to align each of the clinical scan's intensity. Borrowing the idea from external fiducial markers, we expected there might be an appropriate internal reference region relatively unaffected by contrast agent, which would be resilient to post-contrast signal change. In that case, the ratio of the intensity value of post-contrast image to pre-contrast image in the region of interest (ROI) would approach *1*.

The pre-contrast images are segmented using FreeSurfer (Bruce Fischl et al., 2002; Van Leemput et al., 2009) with the hippocampal subfield segmentation applied. All scans were quality checked by a trained reviewer. The post-contrast image was co-registered to the pre-contrast image through intra-subject linear co-registration using FSL FLIRT with six degrees of freedom and correlation ratio cost function (Jenkinson et al., 2002). Although both scans are of the same sequence parameters, we considered the presence of contrast enhancement to require a cost function befitting an inter-modal registration vs. intra-modal registration. The FreeSurfer segmentation masks were also resliced to the pre-contrast image space. The segmentation masks were applied to the pre-contrast and co-registered post-contrast images. Mean values for post/pre-contrast ratio values within ROIs were computed and used as features for subsequent processing.

Forty-five regions are segmented and labelled. To further refine the results, only regions with at least 100 voxels across scans are regarded robust and valid, which results in 38 regions. As shown in Figure 4-2, every region has increased signal with respect to the application of contrast agent.

Ideally, a region wholly unaffected by contrast uptake could be used to correct scaling. However, there is no consistent region within the brain with unequivocal preservation of signal. As such, we wish to find a region that can be consistently segmented with the lowest signal change e.g. coefficient of variation. WM would be expected to have the lowest based on its known cerebro-vasculature. The coefficient of variation for cerebral WM (both left and right) post/pre-contrast ratio is the smallest of all the regions (Figure 4-2).

4.2.3 Scaling correction factor estimation with tissue class segmentation

Since white matter show consistency across scans, I explored the post/pre-contrast ratio on coarser tissue type segmentations. This provides advantages in computational time and robustness, which are very important for large-scale applications.

The pre-contrast image is first brain extracted using FSL BET (Smith, 2002). The brain extraction image is then segmented into WM, GM, and CSF using FSL FAST (Zhang, Brady, & Smith, 2001). To exclude the possible influence of GM on WM, the WM mask was eroded using a morphological 3-dimensional kernel. The post/pre-contrast ratio for each tissue type is generated the same way as described in the previous section.

The ratios for all the scans in the training set are fitted to a normal distribution separately for each tissue type (or region). For a new test scan, the same processing steps are implemented, generating ratio for each tissue type. The ratios are substituted into the probability distribution functions fitted using the training set, generating normalized likelihood. Each ratio can be used to predict the scaling correction factor (SCF): $SCF_R = \hat{\mu}_R / r_R$, where $\hat{\mu}_R$ denotes the mean ratio value of the tissue R in the training set; r_R denotes the ratio value for tissue R from the new scan.

4.2.4 Retrospective TLE analysis

The pre-contrast image was subtracted from the co-registered, scaling corrected post-contrast image to generate a raw subtracted image. The raw subtracted image was normalized by the mean value of grey matter to generate the relative CBV (rCBV) map.

I measured the rCBV in hippocampus and hippocampal subfields. The difference of contralateral and ipsilateral rCBV was used to study the epilepsy lateralization. In the individual analysis, raw data were first visualized with boxplots, and outliers were excluded outside of 1.5 times the interquartile range (IQR).

4.3 Results

4.3.1 Screening internal reference region

The results of screening using FreeSurfer segmentation are shown in Figure 4-2 and suggest that there is not an unaffected anatomically well-defined region with a ratio distributed around 1 in the FreeSurfer segmentation protocol. The scaling correction factor for each scan was estimated in leave-one-out manner. The lowest mean squared error 0.00014, mean absolute error 0.00888 and standard deviation 0.01193 are achieved with left cerebral WM alone.

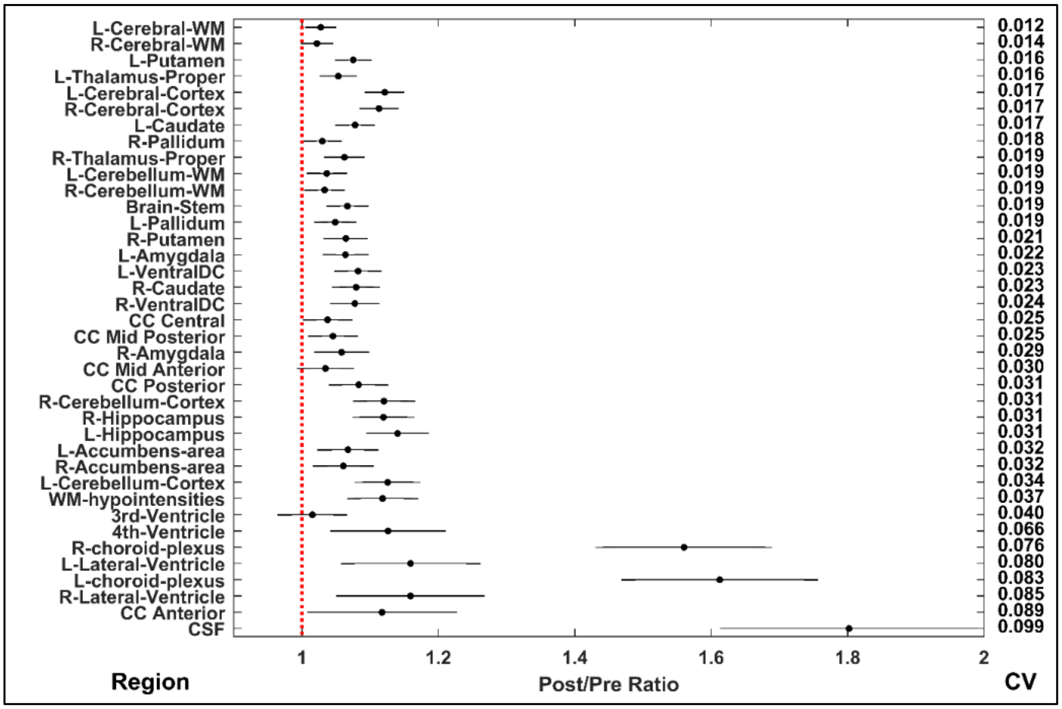


Figure 4-2. Post/pre-contrast ratio values (mean ± std) for regions from FreeSurfer automatic subcortical segmentation pipeline. The regions are sorted in the order of the coefficient of variation (CV) of post/pre-contrast ratio.

For scaling correction using tissue segmentations. Using WM provides the best estimation regarding mean absolute error (0.00998), mean squared error (0.00017); whereas WM/CSF achieves slightly lower standard deviation (0.01292) and similar mean squared error (0.00017). The mean squared error, mean absolute error, and standard deviation levels are close to those achieved using regional segmentations.

4.3.2 Retrospective TLE analysis

All clinical scans were within two standard deviation of the mean of the reference SNR distribution of our existing database. I found a statistically significant relative CBV decrease in the hippocampus ipsilateral to seizure-onset (paired t-test, $N_{\text{Left-TLE}} = 17$, $N_{\text{Right-TLE}} = 6$, $p < 0.05$) with a mean contralateral ipsilateral difference 0.082 (standard deviation, 0.174). This finding is in line with previous studies suggesting ipsilateral hypometabolism in mesial temporal lobe using PET (Sarikaya, 2015; Theodore et al., 1997), ASL (Wolf et al., 2001) and DSC-MRI (Wu et al., 1999).

I found statistically significant mean rCBV change in the ipsilateral subiculum (paired t-test, $N_{\text{Left-TLE}} = 15$, $N_{\text{Right-TLE}} = 6$, $p < 0.05$) with a mean contralateral ipsilateral difference 0.177 (standard deviation, 0.289). The results are trending when examining the left ($p = 0.07$) and right TLE ($p = 0.11$) patients respectively. I found no significant differences in hippocampal volumes or hippocampal subfield volumes (paired t-tests, $p > 0.05$) consistent with a recent large-scale study on brain region volumes in epilepsy without patent lesions (Whelan et al., 2016).

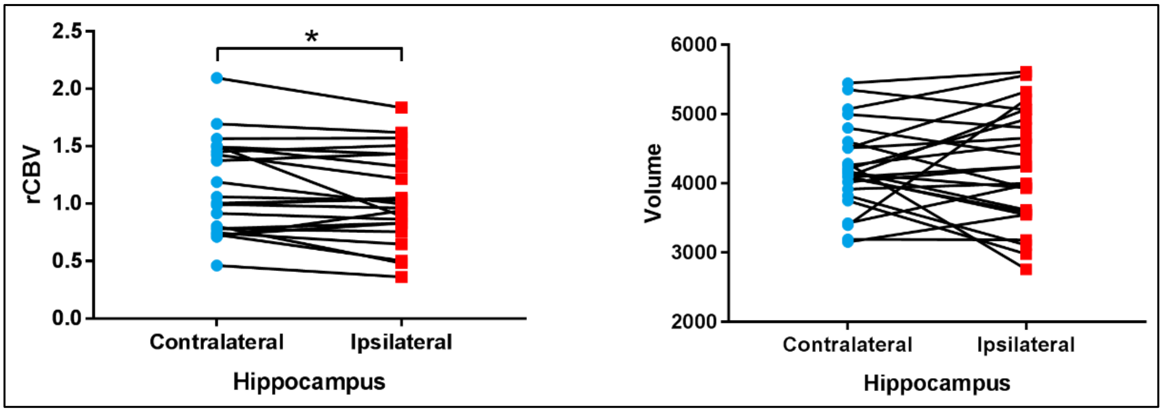


Figure 4-3. The rCBV (left) and volume (right) of hippocampus contralateral and ipsilateral to the seizure onset. There is significant ipsilateral rCBV decrease but not significant structural difference for hippocampus.

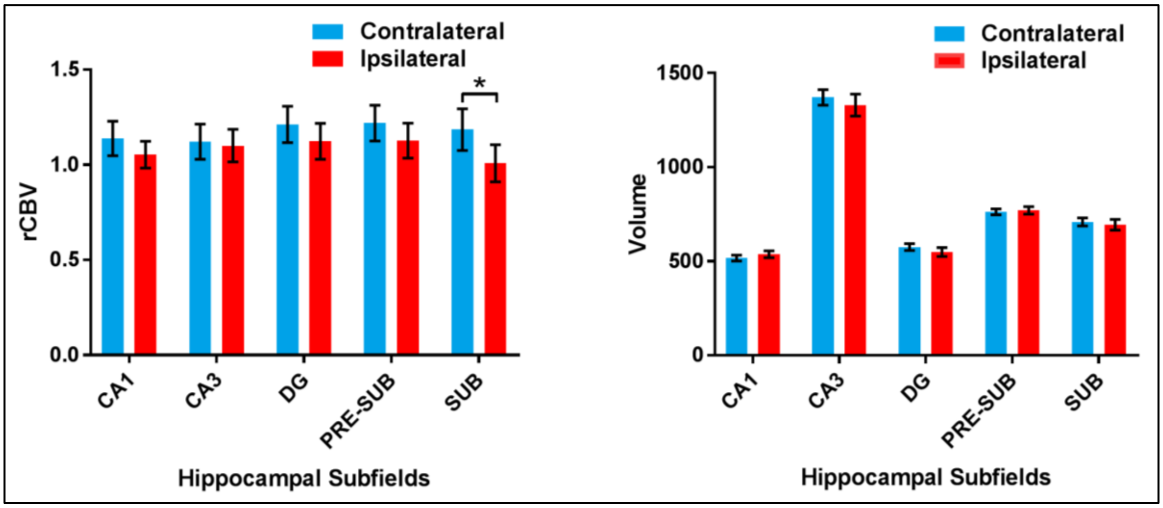


Figure 4-4. The rCBV (left) and volume (right) of hippocampal subfields contralateral and ipsilateral to the seizure onset. There is statistically significant ipsilateral rCBV decrease in subiculum but no significant structural asymmetry for hippocampal subfields. The error-bars show the standard error of the mean.

4.4 Discussion

In the clinical contrast-enhanced MRI scans, I demonstrated that tissue segmentation based methods can achieve accurate scaling correction. WM plays the primary part in the estimation. The use of white matter as a reference region is also supported by results from brain volume measurement (Maclaren et al., 2014). However, in the event this method was to be used in a cohort with suspected WM signal changes like WM lesions or potential demyelinating diseases, this estimation model can incorporate other tissue classes to improve the robustness potentially mediated by non-uniform WM.

Physiologically speaking, the white matter and cerebrospinal fluid should uptake little gadolinium; while grey matter being more densely vascularized should lead to increased contrast uptake. The different uptake within the GM tissue class across regions across scans renders the ratio of GM more variant than WM. Partial volume effects may also contribute to this variance, though would likely affect the entire image uniformly. The segmentation of GM is likely to include regions of both micro and macro vasculature, which reflects the greatest contrast after contrast agent application. Though in principle, CSF uptakes little gadolinium; the segmentation of CSF is also prone to the influence of epicortical vasculature signal change or GM. The mean intensity ratio calculated using the eroded WM masks is similar to the value using the uneroded WM mask, which suggests a relative consistent distribution of gadolinium uptake across the periphery of the white matter and robust segmentation of white matter.

4.4.1 Finer segmentation of white matter

As white matter shows better homogeneity, I further explored the finer structure of cerebral white matter using the FreeSurfer parcellation results. For the WM voxels, the labels are assigned as the label of the closest GM voxel (Salat et al., 2009). As the number of labels is relatively large, the result is not shown here. The coefficients of variation are not significantly smaller than using WM as a whole.

4.4.2 Epilepsy Lateralization

Refractory TLE patients with structurally normal MRI scans often present a challenge to the surgical team with regard to lateralization of seizure onset. Our results suggest that CBV measurements may provide supportive information regarding the laterality of the epileptogenic region. Future studies may explore the role of CBV-fMRI. Additionally, our work suggests the subiculum could be a driver of the metabolic dysfunction, supporting existing research on the subiculum's role in epileptogenesis (Stafstrom, 2005). Electrophysiological studies on extracted hippocampal lesions in TLE have implicated the subiculum as the potential origin of interictal activity (Cohen, Navarro, Clemenceau, Baulac, & Miles, 2002).

4.4.3 Limitations

For inter-subject analysis, it is important to process all of the data in the same pipeline, as there is systematic bias using different methods (Klauschen, Goldman, Barra, Meyer-Lindenberg, & Lundervold, 2009). The validity of the method depends on the stability of the within-method variation. Given the evaluation in publicly available software, FAST and FreeSurfer are likely to provide accurate and reproducible results for this analytical stream (Valverde, Oliver, Cabezas, Roura, & Lladó, 2015). With regards to FreeSurfer segmentations not revealing an unaffected region for contrast, it is possible that other structures (such as maxillary sinus) not defined in the segmentation template used in this study could be potentially accurate internal reference regions.

By using relative CBV, we are assuming the absence of global changes in CBV. Although hemispheric global grey matter CBV changes are unlikely, it will be important to examine those mean values more carefully to ensure that those do not inaccurately affect the results.

4.4.4 Conclusions

In this study, we provide evidence for ipsilateral hippocampal CBV changes in patients with unilateral TLE using intensity corrected steady state contrast enhanced CBV-fMRI. This was achieved with an analytical stream that was agnostic of scanner brand, type and protocol and can be potentially implemented in an informatics system. Since these scans were acquired not for an imaging research study but for potential pre-surgical evaluation, this work demonstrates the potential of the proposed framework to characterize the functional profile of patients who may have retrospective contrast-enhanced scans. This analysis permits generating high-resolution metabolic maps using scans that would already be acquired for clinical evaluation. This applies not only to focal hippocampal metabolic dysfunction, but also to any indication where contrast enhanced MRI is standard or warranted. This analysis and methodological investigation serve as a case study of how novel analytical approaches may serve as a lens to refocus scan protocols to generate clinically relevant imaging measures. Additionally, future studies may focus on hippocampal sub-regional analysis and cognitive correlates, as certain sub-regions have been shown to exhibit selective vulnerability in different disease states.

4.5 Future work

Demonstrating its utility in a TLE lateralization task, the retrospective CBV technique opens up the possibility of standardizing CBV-fMRI from clinical scan database thanks to the large-scale research CBV collection. Future work includes collaboration with clinicians to expand the usage into other neurological conditions including Alzheimer’s disease and normal aging.

Another direction of enriching CBV-fMRI harvesting the current collection of CBV-fMRI scans and large-scale structural MRI scans is to learn a mapping between structural MRI and CBV-fMRI. There are previous works working on synthesizing high-dose contrast-enhanced MRI scans from low-dose contrast-enhanced MRI scans (Gong, Pauly, Wintermark, & Zaharchuk, 2018) for tumor evaluation, but the practicality is still limited, and the utility for more subtle functional changes was not demonstrated. Given the mapping, the structural MRIs can be first transformed to CBV-fMRI space and then first processed. This mapping could potentially improves the diagnosis despite deriving all information from structure because of the search space constraints set by this prior.

The enrichment of CBV-fMRI dataset potentially enables deep learning based analytic approaches on CBV-fMRI with a similar pipeline used in structural MRI.

5 Discussion and conclusion

In this thesis, I developed a study protocol of using large-scale neuroimaging data to study Alzheimer's disease and normal aging from univariate parametric mapping based characterization to advanced machine learning based discrimination. Univariate parametric mapping can be best used for localization-based interpretation, which is important for understanding underlying neuroanatomical and neurophysiological by extracting mechanistic information. Whereas machine learning approach is best at discriminative analysis, which is important for accurate disease diagnosis. Both are important in terms of biomarker development. I further investigated localization through machine learning models, and the linkage of the machine learning derived measures with other biomarkers. Utilizing large-scale datasets, the developed protocol has the potential to advance the neuroscientific understanding and inform the clinical practice of Alzheimer's disease and normal aging.

The protocol set up through this thesis work is able to transfer to other neurological disease, and neuroimaging modalities, some preliminary explorations along this route are demonstrated in the appendix.

6 Related Publications

Journal Papers:

- **Xinyang Feng**, Maria J. Hamberger, Hannah C. Sigmon, Jia Guo, Scott A. Small, Frank A. Provenzano: Temporal lobe epilepsy lateralization using retrospective cerebral blood volume MRI. NeuroImage: Clinical, 2018.
- **Xinyang Feng**, Jie Yang, Zachary C. Lipton, Scott A. Small, Frank A. Provenzano: Deep learning on MRI affirms the prominence of the hippocampal formation in Alzheimer's disease classification. Brain, 2018 (under review, on bioRxiv 456277).
- **Xinyang Feng**, Frank A. Provenzano, Scott A. Small: Diagnosing prodromal Alzheimer's disease with conventional MRI through deep learning. Lancet, 2019 (submitted).
- **Xinyang Feng**, Jia Guo, Hannah C. Sigmon, Richard P. Sloan, Adam M. Brickman, Frank A. Provenzano, Scott A. Small: Brain regions vulnerable and resistant to aging without Alzheimer's disease. PNAS, 2019 (under review).
- Sabrina Simoes, Jia Guo, Luna Buitrago, **Xinyang Feng**, Etty Cortes, Diego E. Berman, Milankumar Kothiya, Vivek Patel, Young-Hyun Kim, Kyu-Tae Chang, S. Abid Hussaini, Herman Moreno, Gilbert Di Paolo, Scott A. Small: Alzheimer's vulnerable brain region relies on a neuronal-distinct recycling endosome. Nature, 2018 (under revision).
- **Xinyang Feng**, Zachary C. Lipton, Jie Yang, Scott A. Small, Frank A. Provenzano: Deep learning based brain age estimation in healthy population using structural MRI. Manuscript prepared, 2019.
- Jie Yang, **Xinyang Feng**, Andrew F. Laine, and Elsa D. Angelini: Characterizing Alzheimer's disease with image and genetic biomarkers using supervised topic models. IEEE Journal of

Biomedical and Health Informatics, 2019 (under minor revision).

- Frank A. Provenzano, Jia Guo, Melanie M Wall, **Xinyang Feng**, Hannah C Sigmon, Gary Brucato, Michael B First, Douglas L Rothman, Ragy R Girgis, Jeffrey A Lieberman, Scott A Small: Hippocampal Pathology in Clinical High-Risk Patients and the Onset of Schizophrenia. Biological Psychiatry, 2019 (under review).
- **Xinyang Feng**, Frank Provenzano, Paul Appelbaum, Michael Masucci, Gary Brucato, Jeffrey Lieberman, Ragy Girgis: Amygdalar Volume and Violent Ideation in a Sample at Clinical High-Risk for Psychosis. Psychiatry Research: Neuroimaging, 2019.

Conference Papers:

- **Xinyang Feng**, Jie Yang, Andrew F. Laine, and Elsa D. Angelini: Discriminative localization in CNNs for weakly-supervised segmentation of pulmonary nodules. In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 568-576. Springer, Cham, 2017.
- **Xinyang Feng**, Jie Yang, Andrew F. Laine, and Elsa D. Angelini: Alzheimer's disease diagnosis based on anatomically stratified texture analysis of the hippocampus in structural MRI. In 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI), pp. 1546-1549. IEEE, 2018.

Conference Abstracts:

- **Xinyang Feng**, Frank A. Provenzano, Marla J. Hamberger, Scott A. Small: Temporal lobe epilepsy lateralization using retrospective contrast enhanced imaging. American Epilepsy Society (AES), Annual Meeting 2017.

- **Xinyang Feng**, Frank A. Provenzano, Scott A. Small: Retrospective Functional MRI: Expanding translational imaging to the clinical domain. Society for Neuroscience (SfN), Annual Meeting 2017.
- Jia Guo, **Xinyang Feng**, Hannah C. Sigmon, Frank A. Provenzano, Scott A. Small. MouseStream: An Open Source Software Suit for Mapping and Analyzing Mouse Cortical Functional Architecture In Vivo Using Magnetic Resonance Microscopy. International Society of Magnetic Resonance in Medicine (ISMRM), Annual Meeting, 2018.

7 Abbreviations

Aβ: β-amyloid

AD: Alzheimer’s disease

ADNC: Alzheimer's disease neuropathologic change score

ADNI: Alzheimer’s Disease Neuroimaging Initiative

AI: artificial intelligence

AIC: Akaike information criterion

ANOVA: Analysis of variance

ANTS: Advanced normalization tools

ApoE: Apolipoprotein E

AUC: area under the curve

AUROC: area under the receiver operating characteristic curve

BFRT: Benton face recognition test

BN: batch normalization

BOLD: blood oxygen level dependent

bvFTD: behavioral variant FTD

CA: Cornu Ammonis areas

CAM: class activation mapping

CBS: Corticobasal syndrome

CBF: cerebral blood flow

CBV: cerebral blood volume

CE-MRI: contrast-enhanced MRI

CERAD: Consortium to establish a registry for Alzheimer's disease score

CN: cognitively normal

CNNs: convolutional neural networks

CNR: contrast-to-noise ratio

CSF: cerebrospinal fluid

CV: coefficient of variation

DG: dentate gyrus

DLMRI: deep learning derived MRI score

DLPET: deep learning derived PET score

DLAV45: deep learning derived AV45-PET score

DLFDG: deep learning derived FDG-PET score

EC: entorhinal cortex

EEG: electroencephalogram

FBB: Florbetaben

FC: fully-connected

FDG: fluorodeoxyglucose

FOV: field of view

FTD: Frontotemporal dementia

GRE: gradient echo

HC: hippocampus

HCP: Human Connectome Project

HP: hippocampal formation

ICV: intracranial volume

IFG: inferior frontal gyrus

IQR: interquartile range

LEC: lateral entorhinal cortex

MAE: mean absolute error

MCI: mild cognitive impairment

MCIp: mild cognitive impairment progression

MCIs: mild cognitive impairment stable

MET: magnetoencephalography

MLSUBCA: molecular layer of subiculum and CA fields

MMSE: mini-mental state examination

MPRAGE: magnetization prepared rapid acquisition gradient echo

MRI: magnetic resonance imaging

MTS: medial temporal sclerosis

PD: Parkinson's disease

PET: positron emission tomography

PIB: Pittsburgh Compound B

PNFA: Progressive non-fluent aphasia, aka non-fluent/agrammatic variant primary progressive aphasia (nfvPPA)

POI: parcellation of interest

PRESUB: presubiculum

PSP: Progressive supranuclear palsy

RAVLT: Rey auditory verbal learning test

ReLU: rectified linear unit

ROC: receiver operating characteristic

ROI: region of interest

RMSE: root mean squared error

SCF: scaling correction factor

SNR: Signal to Noise Ratio

SPECT: Single-photon emission computed tomography

svFTD: Semantic variant FTD, aka semantic variant primary progressive aphasia (svPPA)

SUB: subiculum

SUVR: Standardized uptake value ratio

T1w: T1-weighted

TEC: trans-entorhinal cortex

TFE: Turbo Field Echo

TLE: temporal lobe epilepsy

VBA: voxel-based analysis

8 References

Adler, D. H., Wisse, L. E. M., Ittyerah, R., Pluta, J. B., Ding, S.-L., Xie, L., . . . Yushkevich, P. A. (2018). Characterizing the human hippocampus in aging and Alzheimer's disease using a computational atlas derived from ex vivo MRI and histology. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1801093115

Al Zoubi, O., Ki Wong, C., Kuplicki, R. T., Yeh, H.-w., Mayeli, A., Refai, H., . . . Bodurka, J. (2018). Predicting Age From Brain EEG Signals—A Machine Learning Approach. *Frontiers in Aging Neuroscience, 10*(184). doi:10.3389/fnagi.2018.00184

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., . . . Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia, 7*(3), 270-279. doi:10.1016/j.jalz.2011.03.008

Alzheimer's Association. (2018). 2018 Alzheimer's disease facts and figures. *Alzheimer's & Dementia, 14*(3), 367-429. doi:10.1016/j.jalz.2018.02.001

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage, 38*(1), 95-113.

Augustinack, J. C., Huber, K. E., Stevens, A. A., Roy, M., Frosch, M. P., van der Kouwe, A. J. W., . . . Fischl, B. (2013). Predicting the location of human perirhinal cortex, Brodmann's area 35, from MRI. *Neuroimage, 64*, 32-42. doi:10.1016/j.neuroimage.2012.08.071

Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., & Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage, 54*(3), 2033-2044. doi:10.1016/j.neuroimage.2010.09.025

Beach, T. G., Monsell, S. E., Phillips, L. E., & Kukull, W. (2012). Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005–2010. *Journal of neuropathology and experimental neurology, 71*(4), 266-273.

Braak, H., & Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. *Acta neuropathologica, 82*(4), 239-259.

Brickman, A. M., Khan, U. A., Provenzano, F. A., Yeung, L.-K., Suzuki, W., Schroeter, H., . . . Small, S. A. (2014). Enhancing dentate gyrus function with dietary flavanols improves cognition in older adults. *Nature neuroscience, 17*(12), 1798-1803. doi:10.1038/nn.3850

Brookmeyer, R., Abdalla, N., Kawas, C. H., & Corrada, M. M. (2018). Forecasting the prevalence of preclinical and clinical Alzheimer's disease in the United States. *Alzheimer's & Dementia, 14*(2), 121-129. doi:10.1016/j.jalz.2017.10.009

Burke, S. N., & Barnes, C. A. (2006). Neural plasticity in the ageing brain. *Nature Reviews Neuroscience, 7,* 30. doi:10.1038/nrn1809

Cacciaglia, R., Molinuevo, J. L., Falcón, C., Brugulat-Serrat, A., Sánchez-Benavides, G., Gramunt, N., . . . Gispert, J. D. (2018). Effects of APOE-ε4 allele load on brain morphology in a cohort of middle-aged healthy individuals with enriched genetic risk for Alzheimer's disease. *Alzheimer's & Dementia*. doi:10.1016/j.jalz.2018.01.016

Chang, Y. L., Bondi, M. W., Fennema-Notestine, C., McEvoy, L. K., Hagler, D. J., Jr., Jacobson, M. W., . . . Alzheimer's Disease Neuroimaging, I. (2010). Brain substrates of learning and retention in mild cognitive impairment diagnosis and progression to Alzheimer's disease. *Neuropsychologia, 48*(5), 1237-1247. doi:10.1016/j.neuropsychologia.2009.12.024

Chetelat, G., Landeau, B., Salmon, E., Yakushev, I., Bahri, M. A., Mezenge, F., . . . Fellgiebel, A. (2013). Relationships between brain metabolism decrease in normal aging and changes in structural and functional connectivity. *Neuroimage, 76*(1), 167-177. doi:10.1016/j.neuroimage.2013.03.009

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). *3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation.* Paper presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Cham.

Clark, C. M., Schneider, J. A., Bedell, B. J., Beach, T. G., Bilker, W. B., Mintun, M. A., . . . Group, A.-A. S. (2011). Use of Florbetapir-PET for Imaging β-Amyloid Pathology. *Jama, 305*(3), 275-283. doi:10.1001/jama.2010.2008

Cohen, I., Navarro, V., Clemenceau, S., Baulac, M., & Miles, R. (2002). On the Origin of Interictal Activity in Human Temporal Lobe Epilepsy in Vitro. *Science, 298*(5597), 1418-1421. doi:10.1126/science.1076510

Cole, J. H., Leech, R., & Sharp, D. J. (2015). Prediction of brain age suggests accelerated atrophy after traumatic brain injury. *Annals of neurology, 77*(4), 571-581. doi:10.1002/ana.24367

Cole, J. H., Poudel, R. P. K., Tsagkrasoulis, D., Caan, M. W. A., Steves, C., Spector, T. D., & Montana, G. (2017). Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *Neuroimage*. doi:10.1016/j.neuroimage.2017.07.059

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage, 9*(2), 179-194.

de Leon, M. J., Convit, A., Wolf, O. T., Tarshish, C. Y., DeSanti, S., Rusinek, H., . . . Fowler, J. (2001). Prediction of cognitive decline in normal elderly subjects with 2-[(18)F]fluoro-2-deoxy-D-glucose/positron-emission tomography (FDG/PET). *Proceedings of the National Academy of Sciences, 98*(19), 10966-10971. doi:10.1073/pnas.191044198

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 837-845.

Desikan, R. S., Segonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., . . . Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage, 31*(3), 968-980. doi:10.1016/j.neuroimage.2006.01.021

Ding, Y., Sohn, J. H., Kawczynski, M. G., Trivedi, H., Harnish, R., Jenkins, N. W., . . . Franc, B. L. (2018). A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using 18F-FDG PET of the Brain. *Radiology, 0*(0), 180958. doi:10.1148/radiol.2018180958

Dou, Q., Chen, H., Lequan, Y., Zhao, L., Qin, J., Defeng, W., . . . Heng, P. A. (2016). Automatic Detection of Cerebral Microbleeds from MR Images via 3D Convolutional Neural Networks. *IEEE Transactions on Medical Imaging*. doi:10.1109/TMI.2016.2528129

Ellis, K. A., Bush, A. I., Darby, D., De Fazio, D., Foster, J., Hudson, P., . . . Ames, D. (2009). The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *Int Psychogeriatr, 21*(4), 672-687. doi:10.1017/S1041610209009405

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115-118. doi:10.1038/nature21056

Fagan, A. M., Roe, C. M., Xiong, C., Mintun, M. A., Morris, J. C., & Holtzman, D. M. (2007). Cerebrospinal fluid tau/beta-amyloid(42) ratio as a prediction of cognitive decline in nondemented older adults. *Arch Neurol, 64*(3), 343-349. doi:10.1001/archneur.64.3.noc60123

Fagan, A. M., Xiong, C., Jasielec, M. S., Bateman, R. J., Goate, A. M., Benzinger, T. L. S., . . . Holtzman, D. M. (2014). Longitudinal Change in CSF Biomarkers in Autosomal-Dominant Alzheimer’s Disease. *Science translational medicine, 6*(226), 226ra230-226ra230. doi:10.1126/scitranslmed.3007901

Feng, X., Hamberger, M. J., Sigmon, H. C., Guo, J., Small, S. A., & Provenzano, F. A. (2018). Temporal lobe epilepsy lateralization using retrospective cerebral blood volume MRI. *NeuroImage: Clinical, 19,* 911-917. doi:10.1016/j.nicl.2018.05.012

Feng, X., Yang, J., Laine, A. F., & Angelini, E. D. (2017). *Discriminative Localization in CNNs for Weakly-Supervised Segmentation of Pulmonary Nodules.* Paper presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Cham.

Feng, X., Yang, J., Lipton, Z. C., Small, S. A., & Provenzano, F. A. (2018). Deep Learning on MRI Affirms the Prominence of the Hippocampal Formation in Alzheimer's Disease Classification. *bioRxiv*, 456277. doi:10.1101/456277

Fennema-Notestine, C., Panizzon, M. S., Thompson, W. R., Chen, C.-H., Eyler, L. T., Fischl, B., . . . Jernigan, T. L. (2011). Presence of ApoE ε4 allele associated with thinner frontal cortex in middle age. *Journal of Alzheimer's Disease, 26*(s3), 49-60.

Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences, 97*(20), 11050-11055.

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., . . . Klaveness, S. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron, 33*(3), 341-355.

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. *Neuroimage, 9*(2), 195-207.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D. H., . . . Dale, A. M. (2004). Automatically parcellating the human cerebral cortex. *Cerebral Cortex, 14*(1), 11-22.

Franke, K., Ziegler, G., Klöppel, S., & Gaser, C. (2010). Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: Exploring the influence of various parameters. *Neuroimage, 50*(3), 883-892.

French, L., & Paus, T. (2015). A FreeSurfer view of the cortical transcriptome generated from the Allen Human Brain Atlas. *Frontiers in neuroscience, 9*(323). doi:10.3389/fnins.2015.00323

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., & Frackowiak, R. S. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Human brain mapping, 2*(4), 189-210.

Fritsch, V., Da Mota, B., Loth, E., Varoquaux, G., Banaschewski, T., Barker, G. J., . . . Thirion, B. (2015). Robust regression for large-scale neuroimaging studies. *Neuroimage, 111*, 431-441. doi:10.1016/j.neuroimage.2015.02.048

Gaser, C., Franke, K., Klöppel, S., Koutsouleris, N., Sauer, H., & Initiative, A. s. D. N. (2013). BrainAGE in Mild Cognitive Impaired Patients: Predicting the Conversion to Alzheimer’s Disease. *PLoS ONE, 8*(6), e67346. doi:10.1371/journal.pone.0067346

Gazzaley, A., Cooney, J. W., Rissman, J., & D'Esposito, M. (2005). Top-down suppression deficit underlies working memory impairment in normal aging. *Nature neuroscience, 8*, 1298. doi:10.1038/nn1543

Girshick, R. (2015). *Fast R-CNN*. Paper presented at the IEEE International Conference on Computer Vision (ICCV).

Gómez-Isla, T., Price, J. L., McKeel Jr., D. W., Morris, J. C., Growdon, J. H., & Hyman, B. T. (1996). Profound Loss of Layer II Entorhinal Cortex Neurons Occurs in Very Mild Alzheimer’s

Disease. *The Journal of Neuroscience, 16*(14), 4491-4500. doi:10.1523/jneurosci.16-14-04491.1996

Gong, E., Pauly, J. M., Wintermark, M., & Zaharchuk, G. (2018). Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI. *Journal of Magnetic Resonance Imaging*, n/a-n/a. doi:10.1002/jmri.25970

González, R. G., Fischman, A. J., Guimaraes, A. R., Carr, C. A., Stern, C. E., Halpern, E. F., . . . Rosen, B. R. (1995). Functional MR in the evaluation of dementia: correlation of abnormal dynamic cerebral blood volume measurements with changes in cerebral metabolism on positron emission tomography with fludeoxyglucose F 18. *American journal of neuroradiology, 16*(9), 1763-1770.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*: MIT press.

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., . . . Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama, 316*(22), 2402-2410. doi:10.1001/jama.2016.17216

Gutierrez Becker, B., Klein, T., & Wachinger, C. (2018). Gaussian process uncertainty in age estimation as a measure of brain abnormality. *Neuroimage, 175*, 246-258. doi:10.1016/j.neuroimage.2018.03.075

Hazlett, H. C., Gu, H., Munsell, B. C., Kim, S. H., Styner, M., Wolff, J. J., . . . The, I. N. (2017). Early brain development in infants at high risk for autism spectrum disorder. *Nature, 542*(7641), 348-351. doi:10.1038/nature21369

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*.

Herholz, K., Salmon, E., Perani, D., Baron, J. C., Holthoff, V., Frölich, L., . . . Heiss, W. D. (2002). Discrimination between Alzheimer Dementia and Controls by Automated Analysis of Multicenter FDG PET. *Neuroimage, 17*(1), 302-316. doi:10.1006/nimg.2002.1208

Hoffman, J. M., Welsh-Bohmer, K. A., Hanson, M., Crain, B., Hulette, C., Earl, N., & Coleman, R. E. (2000). FDG PET imaging in patients with pathologically verified dementia. *J Nucl Med, 41*(11), 1920-1928.

Holmes, A. J., Hollinshead, M. O., O'Keefe, T. M., Petrov, V. I., Fariello, G. R., Wald, L. L., . . . Buckner, R. L. (2015). Brain Genomics Superstruct Project initial data release with structural, functional, and behavioral measures. *Scientific data, 2*, 150031. doi:10.1038/sdata.2015.31

Hosseini-Asl, E., Ghazal, M., Mahmoud, A., Aslantas, A., Shalaby, A. M., Casanova, M. F., . . . El-Baz, A. (2018). Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network. *Frontiers in Bioscience (Landmark Ed), 23*, 584-596.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology, 160*(1), 106-154.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology, 195*(1), 215-243.

Hugo G. Schnack, Neeltje E.M. van Haren, Mireille Nieuwenhuis, Hilleke E. Hulshoff Pol, Wiepke Cahn, & René S. Kahn. (2016). Accelerated Brain Aging in Schizophrenia: A Longitudinal Pattern Recognition Study. *American Journal of Psychiatry, 173*(6), 607-616. doi:10.1176/appi.ajp.2015.15070922

Hyman, B. T., Phelps, C. H., Beach, T. G., Bigio, E. H., Cairns, N. J., Carrillo, M. C., . . . Masliah, E. (2012). National Institute on Aging–Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimer's & Dementia, 8*(1), 1-13.

Iglesias, J. E., Augustinack, J. C., Nguyen, K., Player, C. M., Player, A., Wright, M., . . . Initiative, t. A. s. D. N. (2015). A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: Application to adaptive segmentation of in vivo MRI. *Neuroimage, 115*, 117-137.

Illán-Gala, I., Pegueroles, J., Montal, V., Vilaplana, E., Carmona-Iragui, M., Alcolea, D., . . . Fortea, J. (2018). Challenges associated with biomarker-based classification systems for Alzheimer's disease. *Alzheimer's & Dementia, 10*, 346-357. doi:10.1016/j.dadm.2018.03.004

Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., . . . Silverberg, N. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia, 14*(4), 535-562. doi:10.1016/j.jalz.2018.02.018

Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Feldman, H. H., Frisoni, G. B., . . . Dubois, B. (2016). A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology, 87*(5), 539-547. doi:10.1212/wnl.0000000000002923

Jack, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., . . . for the, A. S. (2008). The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI Methods. *Journal of Magnetic Resonance Imaging, 27*(4), 685-691. doi:10.1002/jmri.21049

Jack, Clifford R., & Holtzman, David M. (2013). Biomarker Modeling of Alzheimer’s Disease. *Neuron, 80*(6), 1347-1358. doi:10.1016/j.neuron.2013.12.003

Jack, C. R., Knopman, D. S., Jagust, W. J., Petersen, R. C., Weiner, M. W., Aisen, P. S., . . . Weigand, S. D. (2013). Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *The Lancet Neurology, 12*(2), 207-216.

Jaderberg, M., Simonyan, K., & Zisserman, A. (2015). *Spatial transformer networks.* Paper presented at the Advances in Neural Information Processing Systems (NIPS).

Jagust, W. J., Landau, S. M., Koeppe, R. A., Reiman, E. M., Chen, K., Mathis, C. A., . . . Wang, A. Y. (2015). The Alzheimer's Disease Neuroimaging Initiative 2 PET Core: 2015. *Alzheimers Dement, 11*(7), 757-771. doi:10.1016/j.jalz.2015.05.001

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage, 17*(2), 825-841.

Kalkstein, J., Checksfield, K., Bollinger, J., & Gazzaley, A. (2011). Diminished Top-Down Control Underlies a Visual Imagery Deficit in Normal Aging. *The Journal of Neuroscience, 31*(44), 15768-15774. doi:10.1523/jneurosci.3209-11.2011

Kassem, M. S., Lagopoulos, J., Stait-Gardner, T., Price, W. S., Chohan, T. W., Arnold, J. C., . . . Bennett, M. R. (2013). Stress-Induced Grey Matter Loss Determined by MRI Is Primarily Due to Loss of Dendrites and Their Synapses. *Molecular Neurobiology, 47*(2), 645-661. doi:10.1007/s12035-012-8365-7

Keifer Jr, O. P., Hurt, R. C., Gutman, D. A., Keilholz, S. D., Gourley, S. L., & Ressler, K. J. (2015). Voxel-based morphometry predicts shifts in dendritic spine density and morphology with auditory fear conditioning. *Nature communications, 6*, 7582. doi:10.1038/ncomms8582

Kempermann, G., Kuhn, H. G., & Gage, F. H. (1997). More hippocampal neurons in adult mice living in an enriched environment. *Nature, 386*, 493. doi:10.1038/386493a0

Khan, U. A., Liu, L., Provenzano, F. A., Berman, D. E., Profaci, C. P., Sloan, R., . . . Small, S. A. (2014). Molecular drivers and cortical spread of lateral entorhinal cortex dysfunction in preclinical Alzheimer's disease. *Nature neuroscience, 17*(2), 304-311. doi:10.1038/nn.3606

Killiany, R. J., Hyman, B. T., Gomez-Isla, T., Moss, M. B., Kikinis, R., Jolesz, F., . . . Albert, M. S. (2002). MRI measures of entorhinal cortex vs hippocampus in preclinical AD. *Neurology, 58*(8), 1188-1196. doi:10.1212/wnl.58.8.1188

Klauschen, F., Goldman, A., Barra, V., Meyer‐Lindenberg, A., & Lundervold, A. (2009). Evaluation of automated brain MR image segmentation and volumetry methods. *Human brain mapping, 30*(4), 1310-1327.

Konukoglu, E., Glocker, B., Zikic, D., & Criminisi, A. (2013). Neighbourhood approximation using randomized forests. *Medical image analysis, 17*(7), 790-804. doi:10.1016/j.media.2013.04.013

Korolev, I. O., Symonds, L. L., & Bozoki, A. C. (2016). Predicting Progression from Mild Cognitive Impairment to Alzheimer's Dementia Using Clinical, MRI, and Plasma Biomarkers via Probabilistic Pattern Classification. *PLoS ONE, 11*(2), e0138866. doi:10.1371/journal.pone.0138866

Korolev, S., Safiullin, A., Belyaev, M., & Dodonova, Y. (2017). Residual and Plain Convolutional Neural Networks for 3D Brain MRI Classification. *IEEE International Symposium on Biomedical Imaging (ISBI).*

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *Imagenet classification with deep convolutional neural networks.* Paper presented at the Advances in Neural Information Processing Systems (NIPS).

Lancaster, J., Lorenz, R., Leech, R., & Cole, J. H. (2018). Bayesian Optimization for Neuroimaging Pre-processing in Brain Age Classification and Prediction. *Frontiers in Aging Neuroscience, 10*(28). doi:10.3389/fnagi.2018.00028

Landau, S., & Jagust, W. (2015). Florbetapir processing methods.

Landau, S. M., Harvey, D., Madison, C. M., Koeppe, R. A., Reiman, E. M., Foster, N. L., . . . Jagust, W. J. (2011). Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI. *Neurobiology of Aging, 32*(7), 1207-1218. doi:10.1016/j.neurobiolaging.2009.07.002

Landau, S. M., Harvey, D., Madison, C. M., Reiman, E. M., Foster, N. L., Aisen, P. S., . . . Jagust, W. J. (2010). Comparing predictors of conversion and decline in mild cognitive impairment. *Neurology, 75*(3), 230-238. doi:10.1212/WNL.0b013e3181e8e8b8

Landin-Romero, R., Tan, R., Hodges, J. R., & Kumfor, F. (2016). An update on semantic dementia: genetics, imaging, and pathology. *Alzheimer's Research & Therapy, 8*(1), 52. doi:10.1186/s13195-016-0219-5

Le, T. T., Kuplicki, R. T., McKinney, B. A., Yeh, H.-W., Thompson, W. K., Paulus, M. P., & Tulsa, I. (2018). A Nonlinear Simulation Framework Supports Adjusting for Age When Analyzing BrainAGE. *Frontiers in Aging Neuroscience, 10*, 317-317. doi:10.3389/fnagi.2018.00317

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444. doi:10.1038/nature14539

LeCun, Y., Boser, B. E., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. E., & Jackel, L. D. (1990). *Handwritten digit recognition with a back-propagation network.* Paper presented at the Advances in Neural Information Processing Systems (NIPS).

Lewandowski, N. M., Bordelon, Y., Brickman, A. M., Angulo, S., Khan, U., Muraskin, J., . . . Moreno, H. (2013). Regional vulnerability in Huntington's disease: fMRI-guided molecular analysis in patients and a mouse model of disease. *Neurobiology of disease, 52*, 84-93.

Li, H., Satterthwaite, T. D., & Fan, Y. (2018). *Brain age prediction based on resting-state functional connectivity patterns using convolutional neural networks.* Paper presented at the International Symposium on Biomedical Imaging (ISBI).

Li, Z., Okamoto, K.-I., Hayashi, Y., & Sheng, M. (2004). The Importance of Dendritic Mitochondria in the Morphogenesis and Plasticity of Spines and Synapses. *Cell, 119*(6), 873-887. doi:10.1016/j.cell.2004.11.003

Liew, S.-L., Anglin, J. M., Banks, N. W., Sondag, M., Ito, K. L., Kim, H., . . . Stroud, A. (2018). A large, open source dataset of stroke anatomical brain images and manual lesion segmentations. *Scientific data, 5*, 180011. doi:10.1038/sdata.2018.11

Lin, W., Celik, A., & Paczynski, R. P. (1999). Regional cerebral blood volume: a comparison of the dynamic imaging and the steady state methods. *Journal of Magnetic Resonance Imaging, 9*(1), 44-52.

Lipton, Z. C. (2018). The mythos of model interpretability. *Communications of the ACM (CACM)*.

Liu, W., Wei, D., Chen, Q., Yang, W., Meng, J., Wu, G., . . . Qiu, J. (2017). Longitudinal test-retest neuroimaging data from healthy young adults in southwest China. *Scientific data, 4*, 170017. doi:10.1038/sdata.2017.17

Lohrke, J., Frenzel, T., Endrikat, J., Alves, F. C., Grist, T. M., Law, M., . . . Pietsch, H. (2016). 25 Years of Contrast-Enhanced MRI: Developments, Current Challenges and Future Perspectives. *Advances in therapy, 33*(1), 1-28. doi:10.1007/s12325-015-0275-4

Long, J., Shelhamer, E., & Darrell, T. (2015). *Fully convolutional networks for semantic segmentation.* Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Maclaren, J., Han, Z., Vos, S. B., Fischbein, N., & Bammer, R. (2014). Reliability of brain volume measurements: A test-retest dataset. *Scientific data, 1*.

Malone, I. B., Cash, D., Ridgway, G. R., MacManus, D. G., Ourselin, S., Fox, N. C., & Schott, J. M. (2013). MIRIAD—Public release of a multiple time point Alzheimer's MR imaging dataset. *Neuroimage, 70*, 33-36. doi:10.1016/j.neuroimage.2012.12.044

Marcus, D. S., Fotenos, A. F., Csernansky, J. G., Morris, J. C., & Buckner, R. L. (2010). Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults. *Journal of cognitive neuroscience, 22*(12), 2677-2684.

Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., & Buckner, R. L. (2007). Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *Journal of cognitive neuroscience, 19*(9), 1498-1507.

Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., . . . Taylor, P. (2011). The Parkinson Progression Marker Initiative (PPMI). *Progress in Neurobiology, 95*(4), 629-635. doi:10.1016/j.pneurobio.2011.09.005

Marquié, M., Normandin, M. D., Vanderburg, C. R., Costantino, I. M., Bien, E. A., Rycyna, L. G., . . . Gómez-Isla, T. (2015). Validating novel tau positron emission tomography tracer [F-18]-AV-1451 (T807) on postmortem brain tissue. *Annals of neurology, 78*(5), 787-800. doi:10.1002/ana.24517

Martin, A. J., Friston, K. J., Colebatch, J. G., & Frackowiak, R. S. (1991). Decreases in regional cerebral blood flow with normal aging. *Journal of Cerebral Blood Flow & Metabolism, 11*(4), 684-689. doi:10.1038/jcbfm.1991.121

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., . . . Leemput, K. V. (2015). The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging, 34*(10), 1993-2024. doi:10.1109/TMI.2014.2377694

Miller, M. I., Ratnanather, J. T., Tward, D. J., Brown, T. A., Lee, D. S., Ketcha, M., . . . Team, B. R. (2015). Network Neurodegeneration in Alzheimer’s Disease via MRI Based Shape Diffeomorphometry and High-Field Atlasing. *Frontiers in Bioengineering and Biotechnology, 3,* 54. doi:10.3389/fbioe.2015.00054

Milletari, F., Navab, N., & Ahmadi, S. A. (2016, 25-28 Oct. 2016). *V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation.* Paper presented at the International Conference on 3D Vision (3DV).

Mirra, S. S., Heyman, A., McKeel, D., Sumi, S. M., Crain, B. J., Brownlee, L. M., . . . Berg, L. (1991). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). *Part II. Standardization of the neuropathologic assessment of Alzheimer's disease, 41*(4), 479-479. doi:10.1212/wnl.41.4.479

Mitchell, A. J., & Shiri-Feshki, M. (2009). Rate of progression of mild cognitive impairment to dementia – meta-analysis of 41 robust inception cohort studies. *Acta Psychiatrica Scandinavica, 119*(4), 252-265. doi:10.1111/j.1600-0447.2008.01326.x

Moreno, H., Wu, W. E., Lee, T., Brickman, A., Mayeux, R., Brown, T. R., & Small, S. A. (2007). Imaging the Aβ-related neurotoxicity of Alzheimer disease. *Archives of Neurology, 64*(10), 1467-1477. doi:10.1001/archneur.64.10.1467

Morrison, J. H., & Hof, P. R. (1997). Life and death of neurons in the aging brain. *Science, 278*(5337), 412-419.

Mueller, S. G., Laxer, K. D., Barakos, J., Cheong, I., Garcia, P., & Weiner, M. W. (2009). Subfield atrophy pattern in temporal lobe epilepsy with and without mesial sclerosis detected by high‐resolution MRI at 4 Tesla: Preliminary results. *Epilepsia, 50*(6), 1474-1483.

Mwangi, B., Hasan, K. M., & Soares, J. C. (2013). Prediction of individual subject's age across the human lifespan using diffusion tensor imaging: A machine learning approach. *Neuroimage, 75*(0), 58-67. doi:10.1016/j.neuroimage.2013.02.055

Napper, R. M. A. (2018). Total Number Is Important: Using the Disector Method in Design-Based Stereology to Understand the Structure of the Rodent Brain. *Frontiers in Neuroanatomy, 12*(16). doi:10.3389/fnana.2018.00016

Olsson, B., Lautner, R., Andreasson, U., Öhrfelt, A., Portelius, E., Bjerke, M., . . . Zetterberg, H. (2016). CSF and blood biomarkers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. *The Lancet Neurology, 15*(7), 673-684. doi:10.1016/S1474-4422(16)00070-3

Ossenkoppele, R., Schonhaut, D. R., Schöll, M., Lockhart, S. N., Ayakta, N., Baker, S. L., . . . Rabinovici, G. D. (2016). Tau PET patterns mirror clinical and neuroanatomical variability in Alzheimer's disease. *Brain, 139*(5), 1551-1567. doi:10.1093/brain/aww027

Palmqvist, S., Schöll, M., Strandberg, O., Mattsson, N., Stomrud, E., Zetterberg, H., . . . Hansson, O. (2017). Earliest accumulation of β-amyloid occurs within the default-mode network and concurrently affects brain connectivity. *Nature communications, 8*(1), 1214. doi:10.1038/s41467-017-01150-x

Patterson, C. (2018). *World Alzheimer Report 2018—The state of the art of dementia research: New frontiers*. Retrieved from

Pavlopoulos, E., Jones, S., Kosmidis, S., Close, M., Kim, C., Kovalerchik, O., . . . Kandel, E. R. (2013). Molecular mechanism for age-related memory loss: the histone-binding protein RbAp48. *Science translational medicine, 5*(200), 200ra115-200ra115.

Payan, A., & Montana, G. (2015). *Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks.* Paper presented at the International Conference on Pattern Recognition Applications and Methods (ICPRAM).

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild Cognitive Impairment: Clinical Characterization and Outcome. *Archives of Neurology, 56*(3), 303-308. doi:10.1001/archneur.56.3.303

Poppenk, J., Evensmoen, H. R., Moscovitch, M., & Nadel, L. (2013). Long-axis specialization of the human hippocampus. *Trends in Cognitive Sciences, 17*(5), 230-240. doi:10.1016/j.tics.2013.03.005

Power, M. C., Mormino, E., Soldan, A., James, B. D., Yu, L., Armstrong, N. M., . . . Schneider, J. (2018). Combined neuropathological pathways account for age-related risk of dementia. *Annals of neurology, 84*(1), 10-22. doi:doi:10.1002/ana.25246

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC bioinformatics, 12*(1), 77.

Rodrigue, K., Kennedy, K., Devous, M., Rieck, J., Hebrank, A., Diaz-Arrastia, R., . . . Park, D. (2012). β-Amyloid burden in healthy aging: regional distribution and cognitive consequences. *Neurology, 78*(6), 387-395.

Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-net: Convolutional networks for biomedical image segmentation.* Paper presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention.

Sachdev, P. S., Lipnicki, D. M., Kochan, N. A., Crawford, J. D., Thalamuthu, A., Andrews, G., . . . Consortium, C. S. o. M. i. a. I. (2015). The Prevalence of Mild Cognitive Impairment in Diverse Geographical and Ethnocultural Regions: The COSMIC Collaboration. *PLoS ONE*, *10*(11), e0142388. doi:10.1371/journal.pone.0142388

Salat, D. H., Greve, D. N., Pacheco, J. L., Quinn, B. T., Helmer, K. G., Buckner, R. L., & Fischl, B. (2009). Regional white matter volume differences in nondemented aging and Alzheimer's disease. *Neuroimage, 44*(4), 1247-1258. doi:10.1016/j.neuroimage.2008.10.030

Sarikaya, I. (2015). PET studies in epilepsy. *American Journal of Nuclear Medicine and Molecular Imaging, 5*(5), 416-430.

Scheltens, P., Blennow, K., Breteler, M. M., de Strooper, B., Frisoni, G. B., Salloway, S., & Van der Flier, W. M. (2016). Alzheimer's disease. *Lancet, 388*(10043), 505-517. doi:10.1016/s0140-6736(15)01124-1

Schobel, S. A., Chaudhury, N. H., Khan, U. A., Paniagua, B., Styner, M. A., Asllani, I., . . . Moore, H. (2013). Imaging patients with psychosis and a mouse model establishes a spreading pattern of hippocampal dysfunction and implicates glutamate as a driver. *Neuron, 78*(1), 81-93.

Ségonne, F., Dale, A. M., Busa, E., Glessner, M., Salat, D., Hahn, H. K., & Fischl, B. (2004). A hybrid approach to the skull stripping problem in MRI. *Neuroimage*, *22*(3), 1060-1075. doi:10.1016/j.neuroimage.2004.03.032

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). *Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization*. Paper presented at the IEEE International Conference on Computer Vision (ICCV).

Shamchi, S. P., Khosravi, M., Taghvaei, R., Emamzadehfard, S., Paydary, K., Raynor, W., . . . Werner, T. (2017). Alteration of normal regional brain FDG uptake in normal aging. *Journal of Nuclear Medicine, 58*(supplement 1), 483-483.

Shaw, L. M., Figurski, M., Waligorska, T., & Trojanowski, J. Q. (2016). *An Overview of the first 8 ADNI CSF Batch Analyses*. Retrieved from

Shaw, L. M., Vanderstichele, H., Knapik-Czajka, M., Clark, C. M., Aisen, P. S., Petersen, R. C., . . . Trojanowski, J. Q. (2009). Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Annals of neurology, 65*(4), 403-413. doi:10.1002/ana.21610

Shaw, P., Lerch, J. P., Pruessner, J. C., Taylor, K. N., Rose, A. B., Greenstein, D., . . . Giedd, J. N. (2007). Cortical morphology in children and adolescents with different apolipoprotein E gene polymorphisms: an observational study. *The Lancet Neurology*, *6*(6), 494-500. doi:10.1016/S1474-4422(07)70106-0

Shi, Y., Yamada, K., Liddelow, S. A., Smith, S. T., Zhao, L., Luo, W., . . . Holtzman, D. M. (2017). ApoE4 markedly exacerbates tau-mediated neurodegeneration in a mouse model of tauopathy. *Nature, 549*(7673), 523-527. doi:10.1038/nature24016

Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition.* Paper presented at the International Conference for Learning Representations (ICLR).

Simor, T., Chu, W.-J., Johnson, L., Safranko, A., Doyle, M., Pohost, G. M., & Elgavish, G. A. (1995). In vivo MRI visualization of acute myocardial ischemia and reperfusion in ferrets by the persistent action of the contrast agent Gd (BME-DTTA). *Circulation, 92*(12), 3549-3559.

Sled, J. G., Zijdenbos, A. P., & Evans, A. C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical Imaging, 17*(1), 87-97.

Small, S. A. (2011). Neurobiological correlates of imaging. *Network: Computation in Neural Systems, 22*(1-4), 162-172. doi:10.3109/0954898X.2011.638695

Small, Scott A. (2014). Isolating Pathogenic Mechanisms Embedded within the Hippocampal Circuit through Regional Vulnerability. *Neuron, 84*(1), 32-39. doi:10.1016/j.neuron.2014.08.030

Small, S. A., Chawla, M. K., Buonocore, M., Rapp, P. R., & Barnes, C. A. (2004). Imaging correlates of brain function in monkeys and rats isolates a hippocampal subregion differentially vulnerable to aging. *Proceedings of the National Academy of Sciences, 101*(18), 7181-7186.

Small, S. A., Simoes-Spassov, S., Mayeux, R., & Petsko, G. A. (2017). Endosomal Traffic Jams Represent a Pathogenic Hub and Therapeutic Target in Alzheimer’s Disease. *Trends Neurosci, 40*(10), 592-602. doi:10.1016/j.tins.2017.08.003

Smith, S. M. (2002). Fast robust automated brain extraction. *Human brain mapping, 17*(3), 143-155.

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., . . . Phelps, C. H. (2011). Toward defining the preclinical stages of Alzheimer’s disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia, 7*(3), 280-292. doi:10.1016/j.jalz.2011.03.003

Stafstrom, C. E. (2005). The Role of the Subiculum in Epilepsy and Epileptogenesis. *Epilepsy Currents, 5*(4), 121-129. doi:10.1111/j.1535-7511.2005.00049.x

Steffener, J., Habeck, C., O'Shea, D., Razlighi, Q., Bherer, L., & Stern, Y. (2016). Differences between chronological and brain age are related to education and self-reported physical activity. *Neurobiology of Aging, 40*, 138-144. doi:10.1016/j.neurobiolaging.2016.01.014

Tapiola, T., Alafuzoff, I., Herukka, S. K., Parkkinen, L., Hartikainen, P., Soininen, H., & Pirttila, T. (2009). Cerebrospinal fluid {beta}-amyloid 42 and tau proteins as biomarkers of Alzheimer-type pathologic changes in the brain. *Arch Neurol, 66*(3), 382-389. doi:10.1001/archneurol.2008.596

Tarawneh, R., Head, D., Allison, S., Buckles, V., Fagan, A. M., Ladenson, J. H., . . . Holtzman, D. M. (2015). Cerebrospinal fluid markers of neurodegeneration and rates of brain atrophy in early Alzheimer disease. *JAMA Neurology, 72*(6), 656-665.

Taylor, J. R., Williams, N., Cusack, R., Auer, T., Shafto, M. A., Dixon, M., . . . Henson, R. N. (2017). The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. *Neuroimage, 144,* 262-269. doi:10.1016/j.neuroimage.2015.09.018

Thaker, A. A., Weinberg, B. D., Dillon, W. P., Hess, C. P., Cabral, H. J., Fleischman, D. A., . . . Desikan, R. S. (2017). Entorhinal Cortex: Antemortem Cortical Thickness and Postmortem Neurofibrillary Tangles and Amyloid Pathology. *American journal of neuroradiology, 38*(5), 961-965. doi:10.3174/ajnr.A5133

Thal, D. R., Rüb, U., Orantes, M., & Braak, H. (2002). Phases of Aβ-deposition in the human brain and its relevance for the development of AD. *Neurology, 58*(12), 1791-1800. doi:10.1212/wnl.58.12.1791

Theodore, W. H., Sato, S., Kufta, C. V., Gaillard, W. D., & Kelley, K. (1997). FDG-positron emission tomography and invasive EEG: seizure focus detection and surgical outcome. *Epilepsia, 38*(1), 81-86.

Therneau, T. M., & Grambsch, P. M. (2013). *Modeling survival data: extending the Cox model*: Springer Science & Business Media.

Valizadeh, S. A., Hänggi, J., Mérillat, S., & Jäncke, L. (2017). Age prediction on the basis of brain anatomical measures. *Human brain mapping, 38*(2), 997-1008. doi:10.1002/hbm.23434

Valverde, S., Oliver, A., Cabezas, M., Roura, E., & Lladó, X. (2015). Comparison of 10 brain tissue segmentation methods using revisited IBSR annotations. *Journal of Magnetic Resonance Imaging, 41*(1), 93-101.

van Erp, T., Hibar, D., Rasmussen, J., Glahn, D., Pearlson, G., Andreassen, O., . . . Dale, A. (2015). Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. *Molecular psychiatry*.

van Erp, T. G. M., Walton, E., Hibar, D. P., Schmaal, L., Jiang, W., Glahn, D. C., . . . Turner, J. A. (2018). Cortical brain abnormalities in 4474 individuals with schizophrenia and 5098 controls via the ENIGMA consortium. *Biological psychiatry*. doi:10.1016/j.biopsych.2018.04.023

Van Essen, D. C., Drury, H. A., Joshi, S., & Miller, M. I. (1998). Functional and structural mapping of human cerebral cortex: solutions are in the surfaces. *Proceedings of the National Academy of Sciences, 95*(3), 788-795.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., & Ugurbil, K. (2013). The WU-Minn Human Connectome Project: An overview. *Neuroimage, 80,* 62-79. doi:10.1016/j.neuroimage.2013.05.041

Van Leemput, K., Bakkour, A., Benner, T., Wiggins, G., Wald, L. L., Augustinack, J., . . . Fischl, B. (2009). Automated segmentation of hippocampal subfields from ultra-high resolution in vivo MRI. *Hippocampus, 19*(6), 549-557. doi:10.1002/hipo.20615

Varikuti, D. P., Genon, S., Sotiras, A., Schwender, H., Hoffstaedter, F., Patil, K. R., . . . Eickhoff, S. B. (2018). Evaluation of non-negative matrix factorization of grey matter in age prediction. *Neuroimage, 173,* 394-410. doi:10.1016/j.neuroimage.2018.03.007

Vemuri, P., Wiste, H. J., Weigand, S. D., Shaw, L. M., Trojanowski, J. Q., Weiner, M. W., . . . Initiative, A. s. D. N. (2009). MRI and CSF biomarkers in normal, MCI, and AD subjects: Predicting future clinical change. *Neurology, 73*(4), 294-301. doi:10.1212/WNL.0b013e3181af79fb

Visser, P. J., Kester, A., Jolles, J., & Verhey, F. (2006). Ten-year risk of dementia in subjects with mild cognitive impairment. *Neurology, 67*(7), 1201-1207. doi:10.1212/01.wnl.0000238517.59286.c5

Vos, S. J., Verhey, F., Frolich, L., Kornhuber, J., Wiltfang, J., Maier, W., . . . Visser, P. J. (2015). Prevalence and prognosis of Alzheimer's disease at the mild cognitive impairment stage. *Brain, 138*(Pt 5), 1327-1338. doi:10.1093/brain/awv029

Walhovd, K. B., Westlye, L. T., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., . . . Fjell, A. M. (2011). Consistent neuroanatomical age-related volume differences across multiple samples. *Neurobiology of Aging, 32*(5), 916-932. doi:10.1016/j.neurobiolaging.2009.05.013

Wang, B., & Pham, T. D. (2011). MRI-based age prediction using hidden Markov models. *Journal of neuroscience methods, 199*(1), 140-145.

Wang, L., Alpert, K. I., Calhoun, V. D., Cobia, D. J., Keator, D. B., King, M. D., . . . Ambite, J. L. (2016). SchizConnect: Mediating neuroimaging databases on schizophrenia and related disorders for large-scale integration. *Neuroimage, 124*(Pt B), 1155-1167. doi:10.1016/j.neuroimage.2015.06.065

Ward, A., Arrighi, H. M., Michels, S., & Cedarbaum, J. M. (2012). Mild cognitive impairment: Disparity of incidence and prevalence estimates. *Alzheimer's & Dementia, 8*(1), 14-21. doi:10.1016/j.jalz.2011.01.002

Wegmayr, V., Aitharaju, S., & Buhmann, J. (2018). *Classification of brain MRI with big data and deep 3D convolutional neural networks.* Paper presented at the SPIE Medical Imaging.

Wei, D., Zhuang, K., Chen, Q., Yang, W., Liu, W., Wang, K., . . . Qiu, J. (2018). Structural and functional MRI from a cross-sectional Southwest University Adult lifespan Dataset (SALD). *bioRxiv,* 177279.

West, M. J. (1999). Stereological methods for estimating the total number of neurons and synapses: issues of precision and bias. *Trends Neurosci, 22*(2), 51-61. doi:10.1016/S0166-2236(98)01362-9

Whelan, C. D., Adams, S., Alhusaini, S., Bargalló, N., Bartolini, E., Bernasconi, A., . . . Carr, S. (2016). *ENIGMA-Epilepsy: Brain volume comparisons between 963 epilepsy cases and 1,358 controls.* Paper presented at the Human brain mapping.

Wolf, R. L., Alsop, D. C., Levy-Reis, I., Meyer, P. T., Maldjian, J. A., Gonzalez-Atavales, J., . . . Detre, J. A. (2001). Detection of mesial temporal lobe hypoperfusion in patients with temporal lobe epilepsy by use of arterial spin labeled perfusion MR imaging. *American journal of neuroradiology, 22*(7), 1334-1341.

Woods, R. P., Mazziotta, J. C., & Cherry, S. R. (1993). MRI-PET registration with automated algorithm. *Journal of computer assisted tomography, 17*, 536-536.

Worker, A., Blain, C., Jarosz, J., Chaudhuri, K. R., Barker, G. J., Williams, S. C. R., . . . Simmons, A. (2014). Cortical Thickness, Surface Area and Volume Measures in Parkinson's Disease, Multiple System Atrophy and Progressive Supranuclear Palsy. *PLoS ONE, 9*(12), e114167. doi:10.1371/journal.pone.0114167

Wu, R. H., Bruening, R., Noachtar, S., Arnold, S., Berchtenbreiter, C., Bartenstein, P., . . . Reiser, M. (1999). MR measurement of regional relative cerebral blood volume in epilepsy. *Journal of Magnetic Resonance Imaging, 9*(3), 435-440. doi:10.1002/(SICI)1522-2586(199903)9:3<435::AID-JMRI11>3.0.CO;2-J

Yang, C., Rangarajan, A., & Ranka, S. (2018). *Visual Explanations From Deep 3D Convolutional Neural Networks for Alzheimer's Disease Classification.* Paper presented at the American Medical Informatics Association Annual Symposium.

Yassa, M. A., Mattfeld, A. T., Stark, S. M., & Stark, C. E. (2011). Age-related memory deficits linked to circuit-specific disruptions in the hippocampus. *Proceedings of the National Academy of Sciences, 108*(21), 8873-8878.

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

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). *Learning Deep Features for Discriminative Localization.* Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Zuo, X.-N., Anderson, J. S., Bellec, P., Birn, R. M., Biswal, B. B., Blautzik, J., . . . Castellanos, F. X. (2014). An open science resource for establishing reliability and reproducibility in functional connectomics. *Scientific data, 1*, 140049.

Zuo, Y., Yang, G., Kwon, E., & Gan, W.-B. (2005). Long-term sensory deprivation prevents dendritic spine loss in primary somatosensory cortex. *Nature, 436*, 261. doi:10.1038/nature03715

A Appendix

A.1 Supplementary figure and tables

A.1.1 Specific aim 1

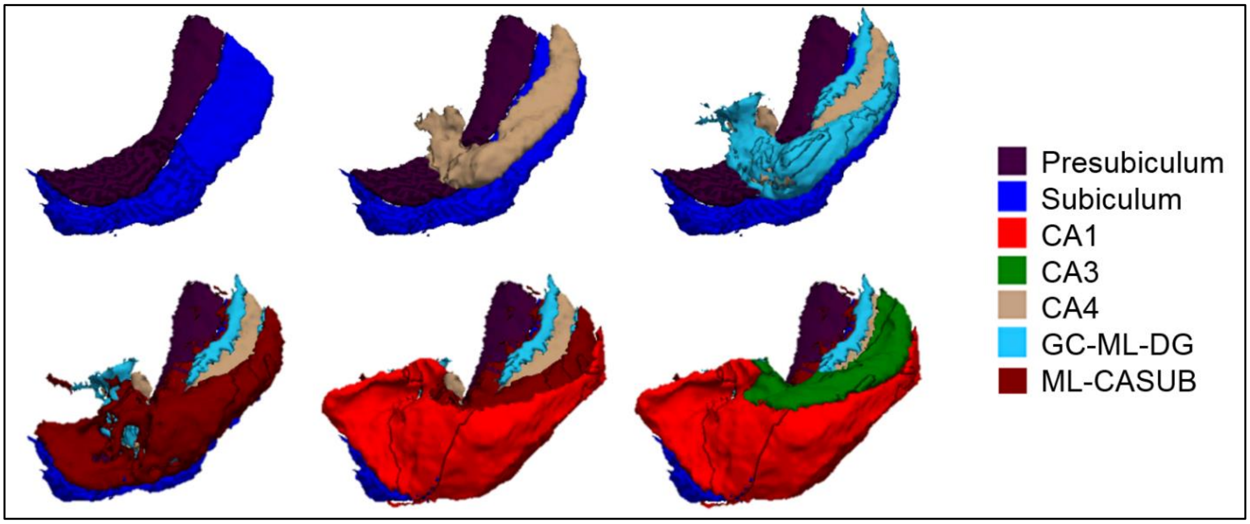


Figure A-1. An illustration of hippocampal subregions.

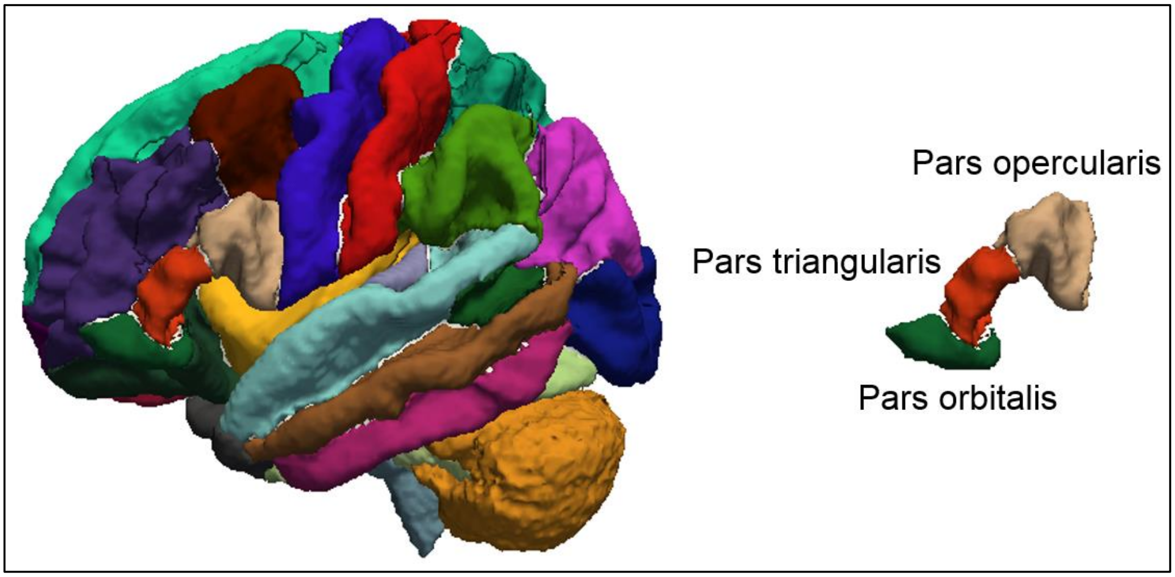


Figure A-2. An illustration of IFG subregions.

A.1.2 Specific aim 2

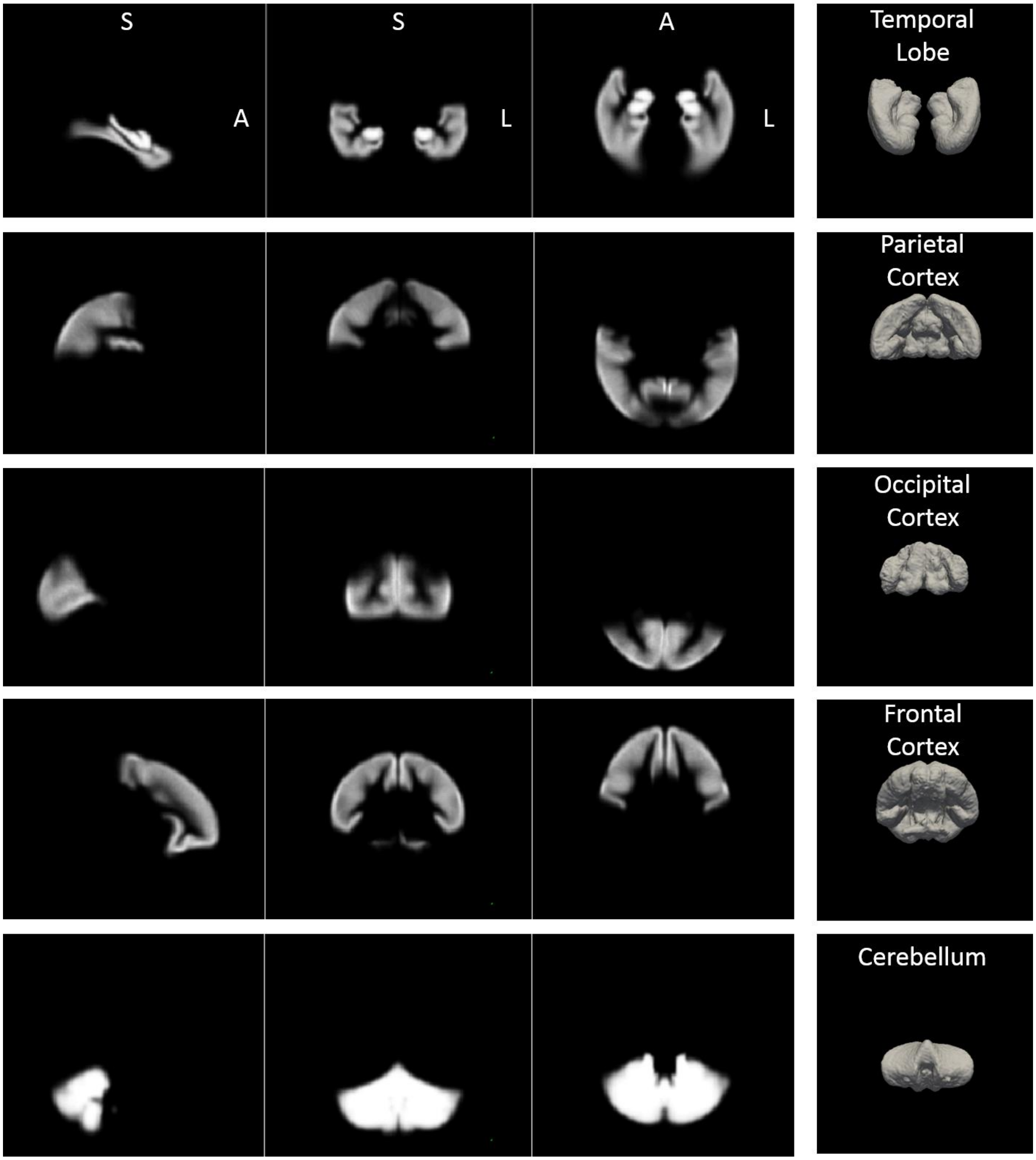
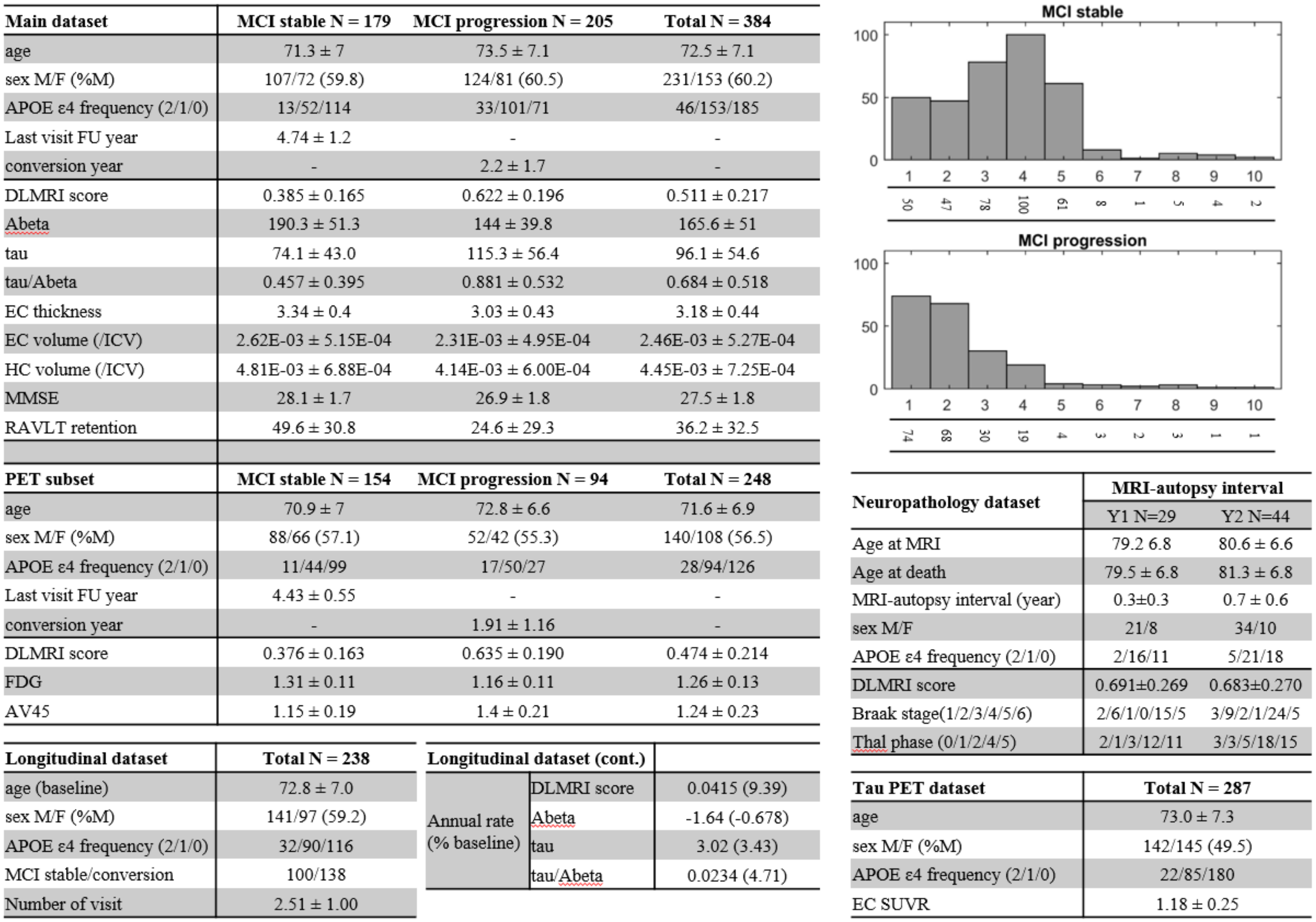


Figure A-3. Brain lobe and cerebellum probability maps.



Main dataset	MCI stable N = 179	MCI progression N = 205	Total N = 384
age	71.3 ± 7	73.5 ± 7.1	72.5 ± 7.1
sex M/F (%M)	107/72 (59.8)	124/81 (60.5)	231/153 (60.2)
APOE ε4 frequency (2/1/0)	13/52/114	33/101/71	46/153/185
Last visit FU year	4.74 ± 1.2	-	-
conversion year	-	2.2 ± 1.7	-
DLMRI score	0.385 ± 0.165	0.622 ± 0.196	0.511 ± 0.217
Abeta	190.3 ± 51.3	144 ± 39.8	165.6 ± 51
tau	74.1 ± 43.0	115.3 ± 56.4	96.1 ± 54.6
tau/Abeta	0.457 ± 0.395	0.881 ± 0.532	0.684 ± 0.518
EC thickness	3.34 ± 0.4	3.03 ± 0.43	3.18 ± 0.44
EC volume (/ICV)	2.62E-03 ± 5.15E-04	2.31E-03 ± 4.95E-04	2.46E-03 ± 5.27E-04
HC volume (/ICV)	4.81E-03 ± 6.88E-04	4.14E-03 ± 6.00E-04	4.45E-03 ± 7.25E-04
MMSE	28.1 ± 1.7	26.9 ± 1.8	27.5 ± 1.8
RAVLT retention	49.6 ± 30.8	24.6 ± 29.3	36.2 ± 32.5

PET subset	MCI stable N = 154	MCI progression N = 94	Total N = 248
age	70.9 ± 7	72.8 ± 6.6	71.6 ± 6.9
sex M/F (%M)	88/66 (57.1)	52/42 (55.3)	140/108 (56.5)
APOE ε4 frequency (2/1/0)	11/44/99	17/50/27	28/94/126
Last visit FU year	4.43 ± 0.55	-	-
conversion year	-	1.91 ± 1.16	-
DLMRI score	0.376 ± 0.163	0.635 ± 0.190	0.474 ± 0.214
FDG	1.31 ± 0.11	1.16 ± 0.11	1.26 ± 0.13
AV45	1.15 ± 0.19	1.4 ± 0.21	1.24 ± 0.23

Longitudinal dataset	Total N = 238
age (baseline)	72.8 ± 7.0
sex M/F (%M)	141/97 (59.2)
APOE ε4 frequency (2/1/0)	32/90/116
MCI stable/conversion	100/138
Number of visit	2.51 ± 1.00

Longitudinal dataset (cont.)		
Annual rate (% baseline)	DLMRI score	0.0415 (9.39)
	Abeta	-1.64 (-0.678)
	tau	3.02 (3.43)
	tau/Abeta	0.0234 (4.71)

Neuropathology dataset	MRI-autopsy interval	
	Y1 N=29	Y2 N=44
Age at MRI	79.2 6.8	80.6 ± 6.6
Age at death	79.5 ± 6.8	81.3 ± 6.8
MRI-autopsy interval (year)	0.3±0.3	0.7 ± 0.6
sex M/F	21/8	34/10
APOE ε4 frequency (2/1/0)	2/16/11	5/21/18
DLMRI score	0.691±0.269	0.683±0.270
Braak stage(1/2/3/4/5/6)	2/6/1/0/15/5	3/9/2/1/24/5
Thal phase (0/1/2/4/5)	2/1/3/12/11	3/3/5/18/15

Tau PET dataset	Total N = 287
age	73.0 ± 7.3
sex M/F (%M)	142/145 (49.5)
APOE ε4 frequency (2/1/0)	22/85/180
EC SUVR	1.18 ± 0.25

Figure A-4. Demographic information and baseline measure of the ADNI MCI cohort.

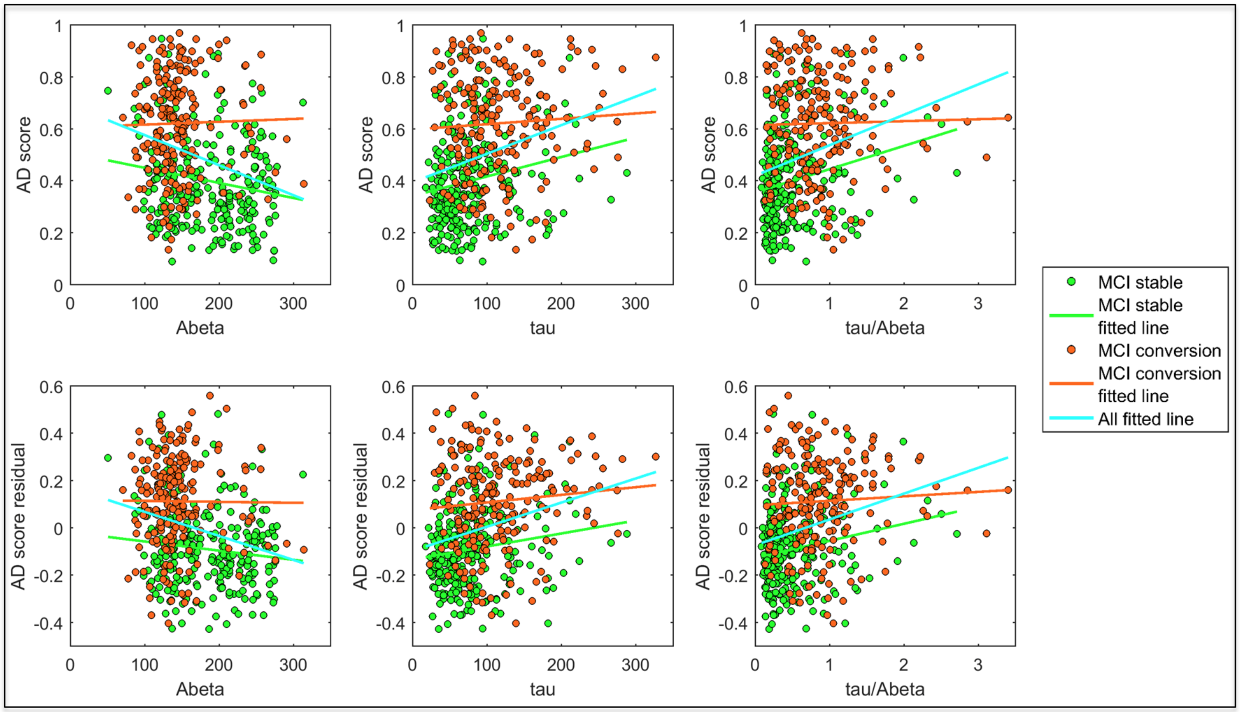


Figure A-5. Correlation of DLMRI score with *in vivo* CSF biomarker measures.

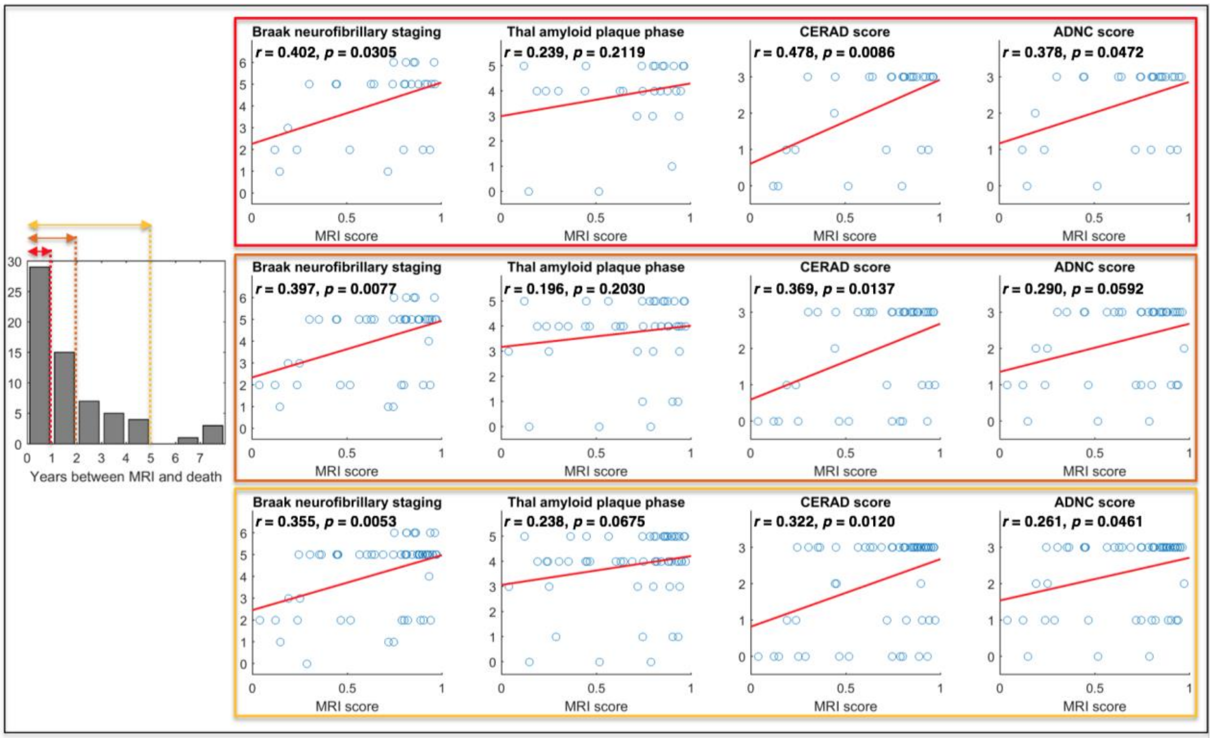


Figure A-6. Correlation of DLMRI score with neuropathological summary measures.

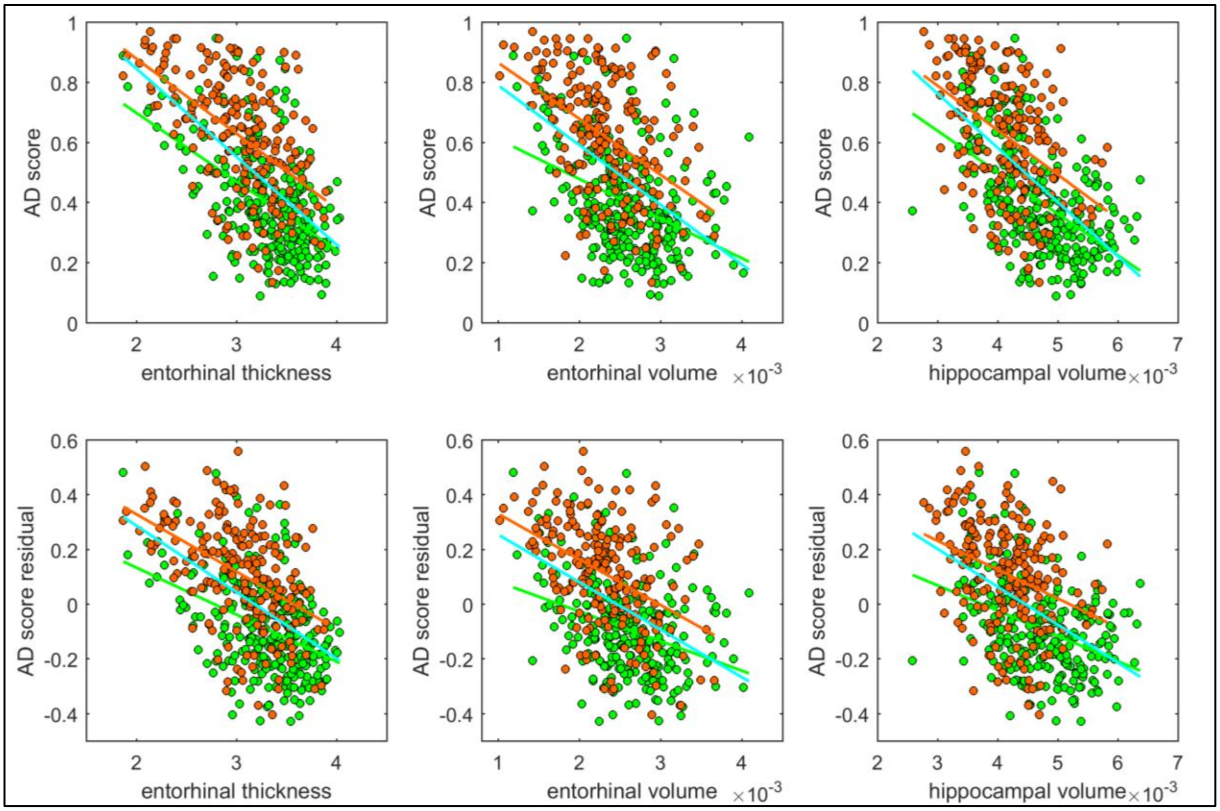


Figure A-7. Correlation of DLMRI score with morphometric measures.

Table A-1. Dataset information of the deep learning age estimation study.

Dataset	Full project name	Source
ADNI	Alzheimer’s Disease Neuroimaging Initiative	http://adni.loni.usc.edu
AIBL	Australian Imaging Biomarkers and Lifestyle Study of Ageing	https://aibl.csiro.au
NIFD	Frontotemporal Lobar Degeneration Neuroimaging Initiative	https://ida.loni.usc.edu/login.jsp?project=NIFD
IXI	Information eXtraction from Images	https://brain-development.org/ixi-dataset
BGSP	Brain Genomics Superstruct Project	https://dataverse.harvard.edu/dataverse/GSP
Cam-CAN	Cambridge Centre for Ageing and Neuroscience	http://www.cam-can.org
OASIS-1	Open Access Series of Imaging Studies-1	https://www.oasis-brains.org
OASIS-2	Open Access Series of Imaging Studies-2	https://www.oasis-brains.org
SALD	Southwest University Adult life-span Dataset	http://fcon_1000.projects.nitrc.org/indi/retro/sald.html
SLIM	Southwest University Longitudinal Imaging	http://fcon_1000.projects.nitrc.org/indi/retro/sout
PPMI	Parkinson’s Progression Markers Initiative	https://www.ppmi-info.org/
SchizConnect	SchizConnect	http://schizconnect.org/
DLBS	Dallas life-span Brain Study	http://fcon_1000.projects.nitrc.org/indi/retro/dlbs.html
CoRR	Consortium for Reliability and Reproducibility	http://fcon_1000.projects.nitrc.org/indi/CoRR/html/

A.2 Cross-sectional volumetric analyses in independent life-span normal aging studies

Following the normal aging study covered in specific aim 2, we also performed cross-sectional volumetric analyses in three independent normal aging studies covering full adult age-span: Cam-CAN (Taylor et al., 2017), IXI, DLBS (Rodrigue et al., 2012). The results show similar patterns with our cohort: dentate gyrus show most reliable age-related decrease within hippocampus, whereas the entorhinal show little age-related difference, IFG show age-related decrease but superior frontal region shows the most reliable age-related decrease. The demographic information of the three datasets are summarized in Table A-1 in section 3.3.

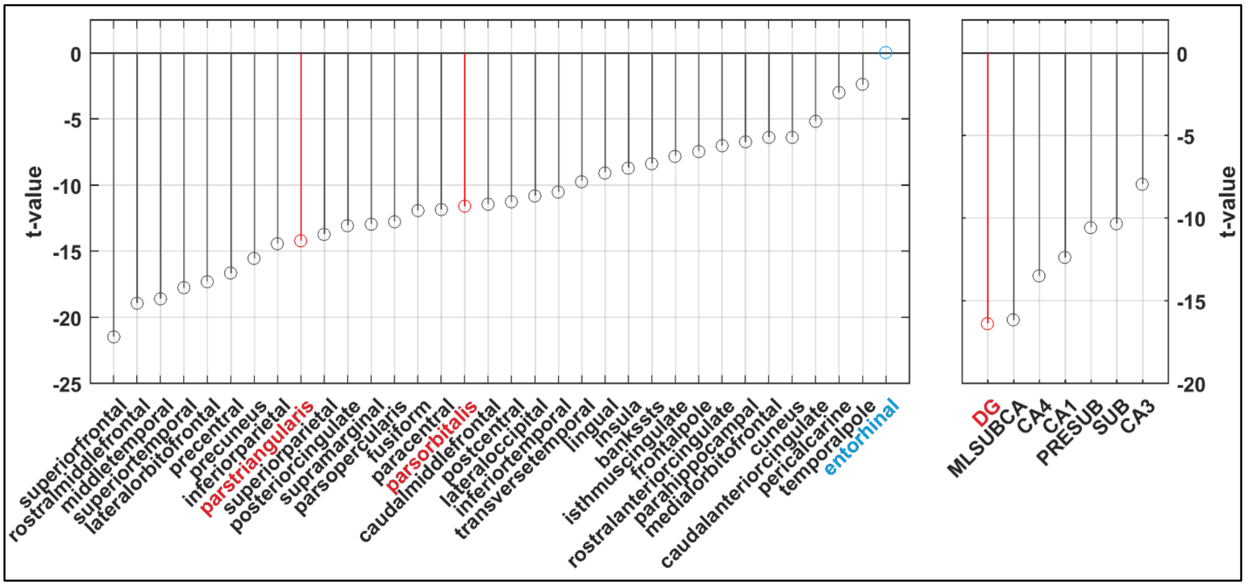


Figure A-8. Age-related changes of left cortical volumes and the hippocampal subfield volumes in Cam-CAN dataset.

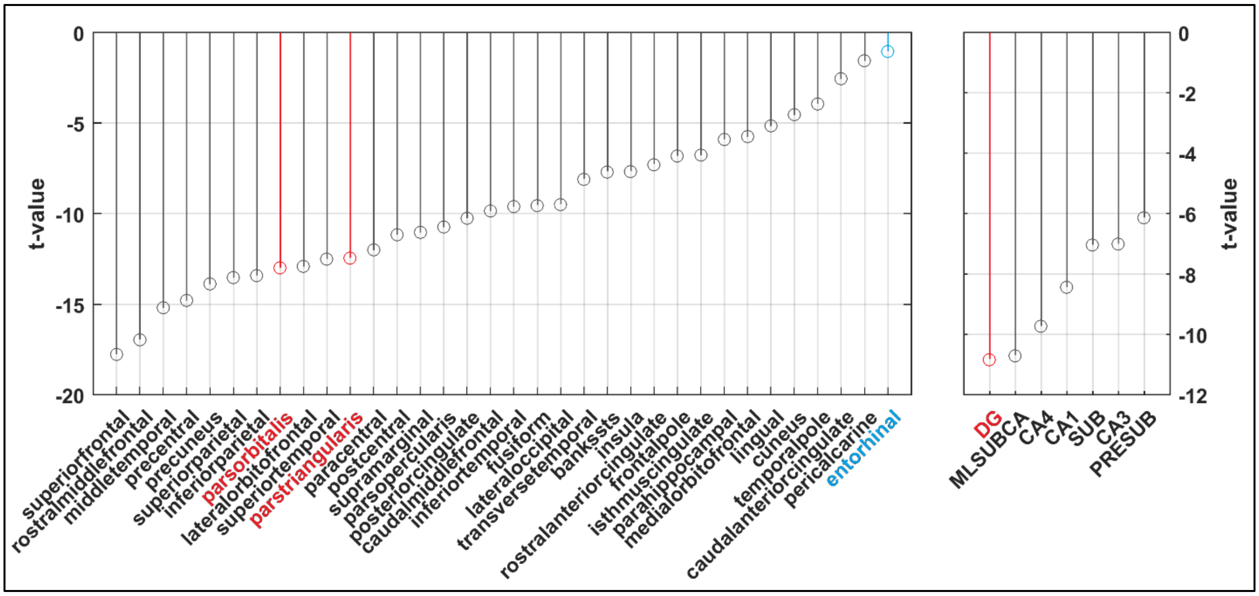


Figure A-9. Age-related changes of left cortical volumes and the hippocampal subfield volumes in DLBS dataset.

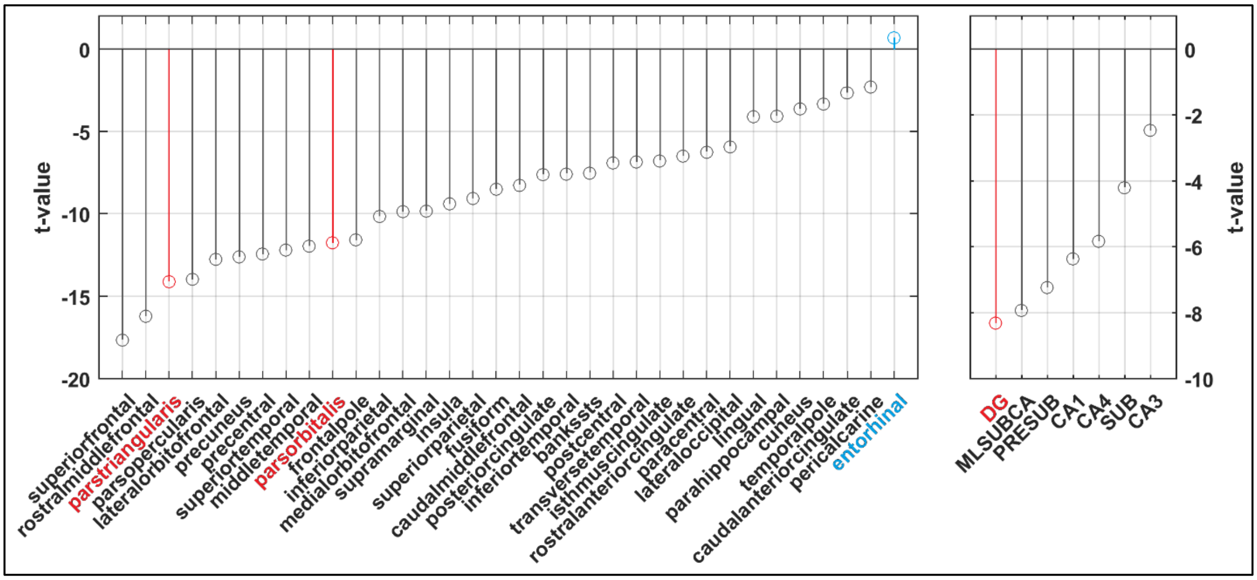


Figure A-10. Age-related changes of the left cortical volumes and the hippocampal subfield volumes in IXI dataset.

A.3 AD classification specificity evaluation

A.3.1 Introduction

In real-world clinical evaluation, we might not be doing binary classification as the settings in the AD classification study. Besides the specificity concept in binary classification, there is a broader concept of specificity in the presence of other neurological conditions. In this section, we evaluate the specificity of the model directly applying the AD model on the data from patients who had other disease types.

In this section, I also present the structural profile of the diseases studied to explain the specificity. The structural profiles are generated in similar ways as the structural profile of AD.

A.3.2 Methods

The specificity was tested on as diverse disease population as possible, including frontotemporal lobe dementia and its different subtypes, Parkinson's disease, and schizophrenia. Additionally, I analyze the structural profile of the different disease to understand the specificity together with what we learned from the interpretation result in AD classification. Specifically, the structural profile is the regional contrast with normal controls.

The FTD dataset is from frontotemporal lobar degeneration neuroimaging initiative and includes the subtypes of FTD: behavior-variant FTD (BV), semantic-variant FTD (SV), progressive nonfluent aphasia (PNFA), corticobasal syndrome (CBS), and progressive supranuclear palsy (PSP). The Parkinson's disease dataset is from PPMI and includes Parkinson's disease, and scans without evidence of dopaminergic deficit (SWEDD) patients. The schizophrenia dataset is from SchizConnect and includes patients with full-blown schizophrenia. Normal control subjects are included in each dataset pertaining to each project's inclusion criteria, and are used to generate the contrast patterns of the individual brain disorders. The basic demographical information is shown in Table A-2.

Table A-2. The demographical information of the data used in the specificity study.

FTD N = 365	Control N = 165	BV N = 61	SV N = 29	PNFA N = 35	CBS N = 45	PSP N = 60
Age	63.3±7.2	61.6±6.6	63.2±6.6	68.6±7.5	66.7±6.8	71.0±7.6
Gender M/F	58/77	40/21	16/13	16/19	21/24	28/32

PD N = 275	Control N = 65	SWEDD N = 41	PD N = 169
Age	60.5±10.7	60.8±10.5	61.1±9.5
Gender M/F	43/22	27/14	108/61

Schiz N = 921	Control N = 470	Schizophrenia N = 451
Age	34.6±12.4	35.8±12.5
Gender M/F	286/184	332/119

A.3.3 Results

A.3.3.1 Specificity

The specificity results are shown in Figure A-11, as boxplots of the probabilistic outputs applying the AD model on the scans from other disease types. Semantic variant FTD show very tight distribution at high probability of being AD, while behavioral variant FTD and PNFA show large variance. The other syndromes generally show good specificity.

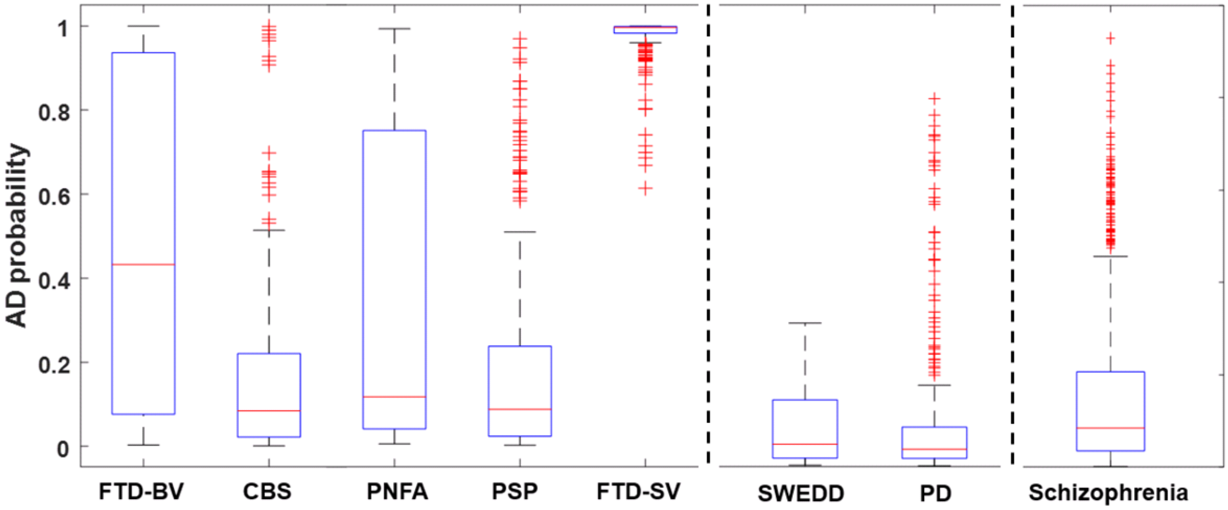


Figure A-11. The specificity of the AD prediction model in the presence of various brain disorders.

A.3.3.2 Structural profile of other neurological diseases

The structural profile of Parkinson's disease is shown in Figure A-12. Patients with Parkinson's disease show little brain atrophy, agreeing with a previous small-scale study (Worker et al., 2014).

The structural profile of FTD including subtypes (SV, BV, PNFA, PSP, CBS) is shown in Figure A-13. Semantic variant FTD is shown in Figure A-14. It's noted that left entorhinal cortex, which overlaps with the region indicated by the class activation map from the AD classification model, is strongly implicated in SV. This explains the low specificity of our model in classifying SV. This structural pattern is consistent with previous studies showing primarily left lateralized atrophy of the anterior temporal lobe regions (Landin-Romero, Tan, Hodges, & Kumfor, 2016).

The structural profile of schizophrenia is shown in Figure A-15. Schizophrenia patients show moderate hippocampal atrophy. The patterns in general agree with previous works (T. van

Erp et al., 2015; T. G. M. van Erp et al., 2018). In addition, this group of schizophrenia patients are in a relative younger age than AD patients.

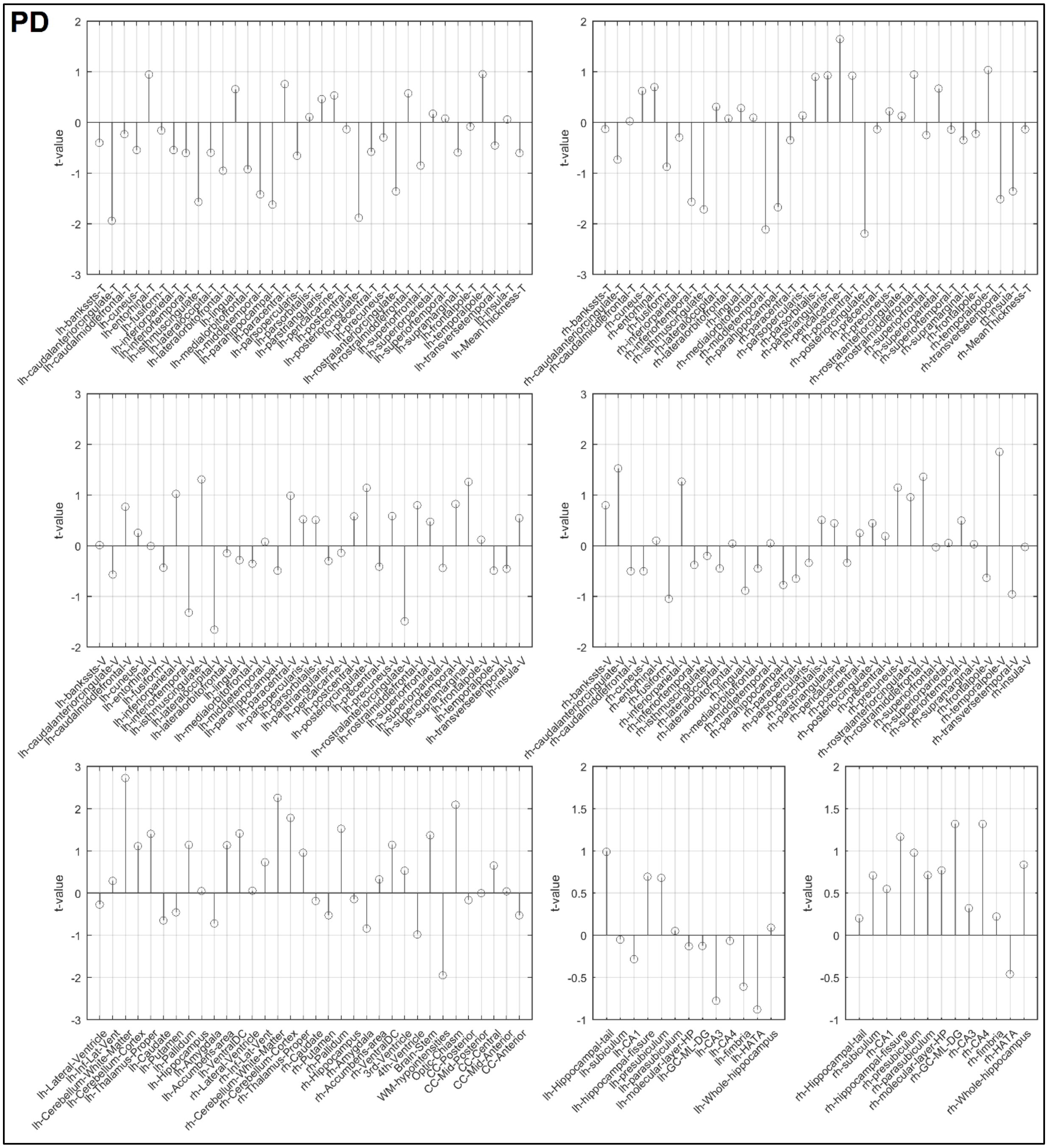


Figure A-12. The structural profile of Parkinson’s disease.

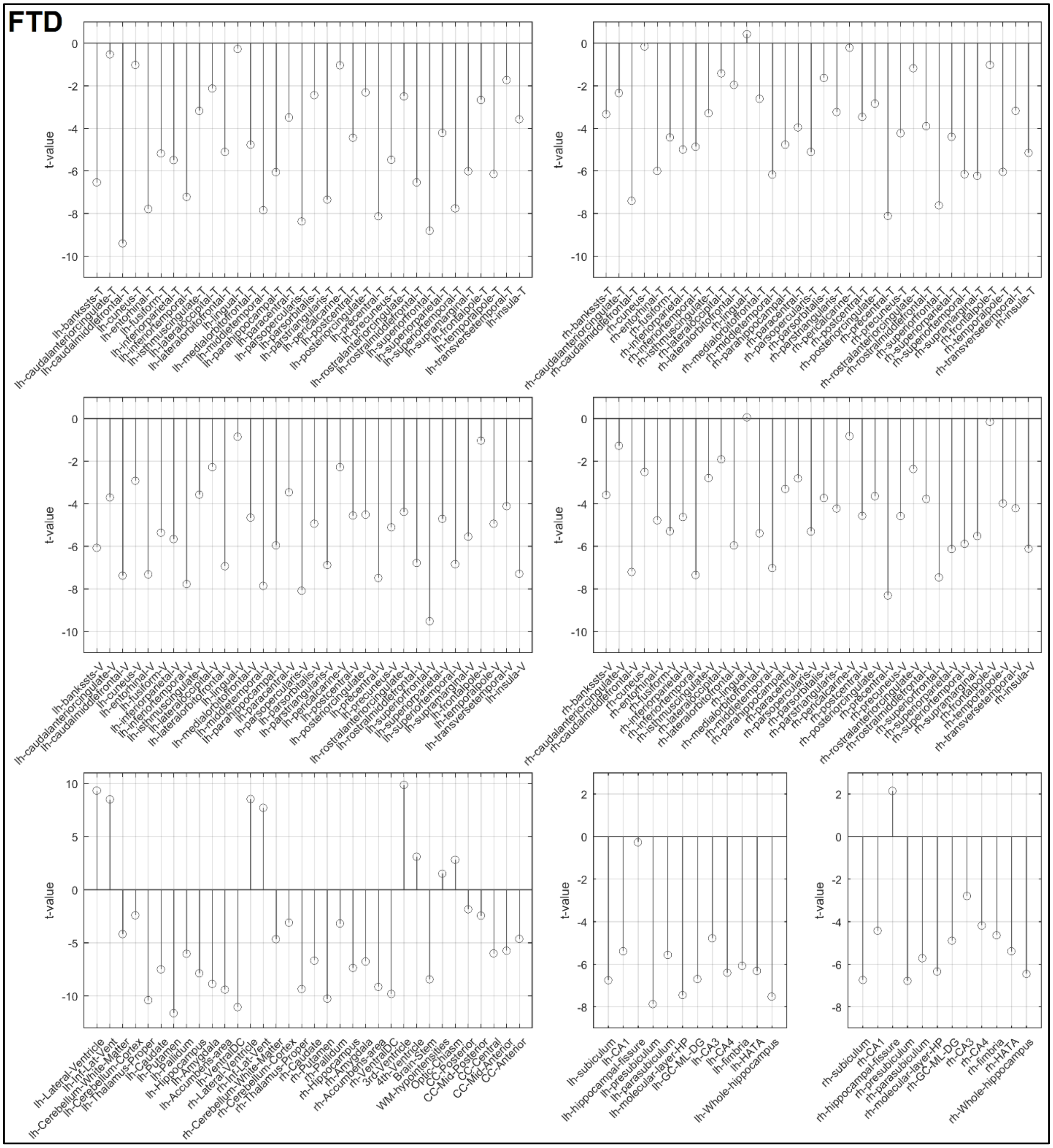


Figure A-13. The structural profile of FTD with all subtypes aggregated.

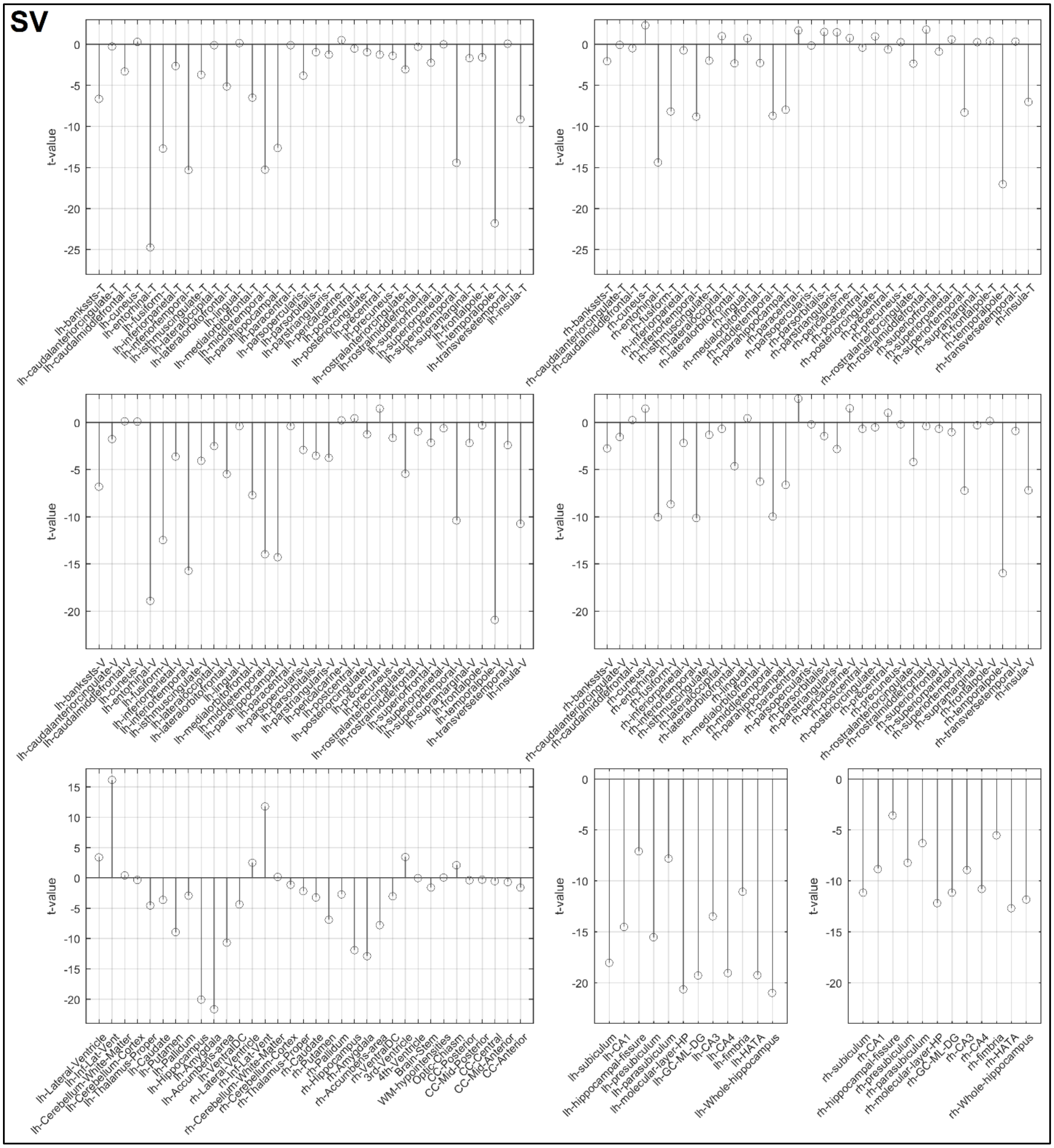


Figure A-14. The structural profile of semantic variant FTD.

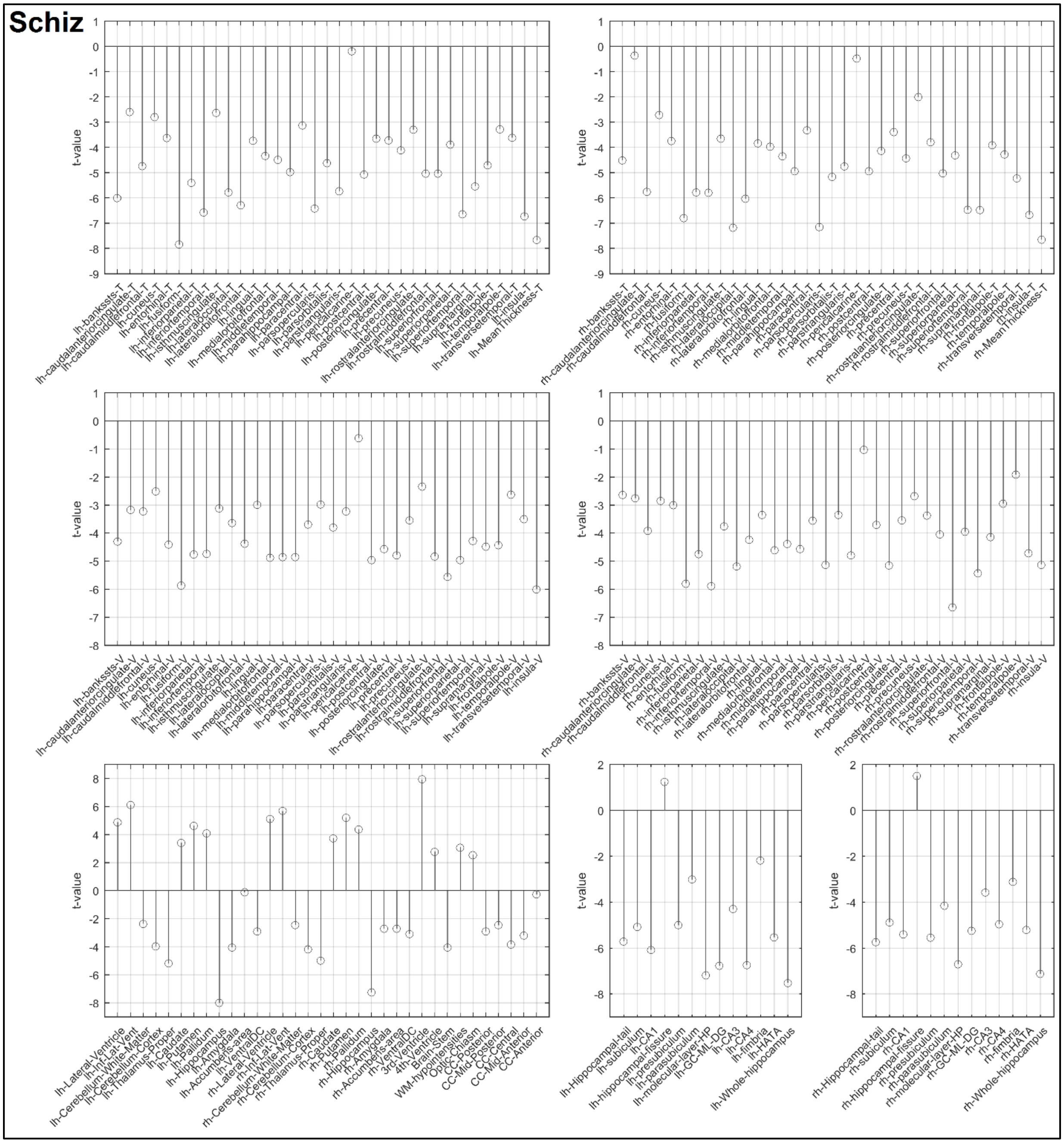


Figure A-15. The structural profile of Schizophrenia.

A.3.4 Discussion

In brain disorders including dementia, the segmentation of atrophic brains is subject to large error. And it is the actual atrophic region that is the hallmark of the diseases. This is less of a problem for machine learning algorithms directly applied on the brain image. The parametric mappings shown in this chapter thus should be interpreted as a reference.

A.4 CSF biomarker estimation from structural MRI

A.4.1 Introduction

In this chapter, I performed a preliminary analysis of estimating *in vivo* CSF biomarker levels from T1-weighted structural MRI, demonstrating a proof-of-concept for a data-driven approach of AD biomarker generation, mixing information from different categories and modalities.

A.4.2 Methods

I included 2954 MRI sessions from 1228 subjects, with valid Aβ and tau measures from CSF at the same visit as the MRI acquisition. The tau levels are log-transformed to alleviate the right-tailed distribution. An illustration of the distribution is shown in Figure A-16. The convolutional neural network structure is the same as the one used in age estimation study.

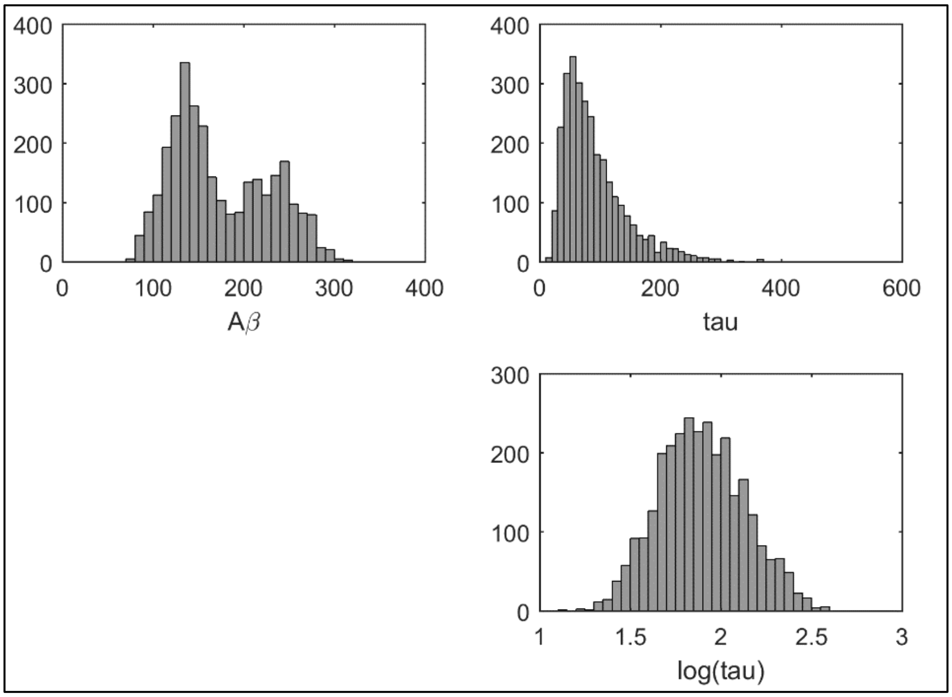


Figure A-16. The distribution of the CSF biomarkers.

A.4.3 Results

The root mean squared error (RMSE) of the Aβ estimation in the test set is 52.4, the correlation coefficient is 0.300 with a p-value 3.09E-7. The RMSE of the tau estimation in the test set is 50.4, the correlation coefficient is 0.416 with a p-value 3.88E-13. The scatterplots and linear fit are shown in Figure A-17.

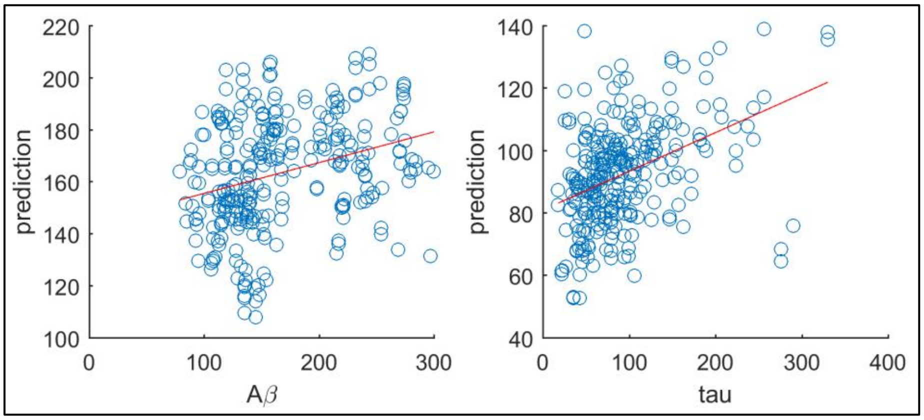


Figure A-17. The predicted CSF biomarker measures versus the ground truth. The red line indicates the linear fit.

A.4.4 Discussion

The current derived measures are still distance away from being used as a surrogate of CSF biomarkers, but they could be useful as correlates of CSF biomarkers, and at the same time reflect structural information as well. Although the correlation is significant showing the derived measure is reflective of the supervision signal, the actual estimated values still have large errors. One primary reason from the observation is the “regression toward the mean” phenomenon, which is a common problem in regression settings, and previously was described in age estimation studies

(Le et al., 2018). There could be potential ways to correct the trend e.g. assigning weights to each sample in proportion to the distance from mean in the training, and is a future direction of the study.

The Aβ estimation analysis showed relatively poor performance comparing with the tau analysis. While the potential methodology limitation cannot be ruled out, this might reflect the temporal sequence of the AD biomarkers as the structural MRI based biomarker is closer to tau biomarker than Aβ biomarker temporally. Or, this could reflect the relationship in the mechanism level: as previously revealed in a tau-PET study that pathological tau aggregation is closely linked to neurodegenerative patterns and clinical manifestations of AD (Ossenkoppele et al., 2016). But this is beyond the power of current evidences.

A.5 Diagnosis and regionality analysis of Alzheimer’s disease using PET

A.5.1 Introduction

In this study, I investigate diagnosing Alzheimer’s disease using PET with deep learning techniques, and perform corresponding regionality analysis. This study follows our previous study using deep learning on structural MRI for AD diagnosis and regionality analysis. Deep learning is known to be a general method incorporating feature engineering and classification in the same model, thus this work is also a proof of concept for the generalizability of the method. I have demonstrated in the previous study (Feng, Yang, et al., 2018), as also described in previous sections of thesis, that using deep learning technique on structural MRI can accurately diagnose Alzheimer’s disease from normal controls, and can predict MCI progression. Moreover, the predictive region identified through the deep learning technique overlaps with our prior knowledge about the pathophysiology of AD. The output probability (referred to as DLMRI score) can be used to summarize the structural information related to AD, and has shown potential clinical utility comparing with other measures.

Despite various practical advantages of structural MRI over other imaging modalities, structural MRI can only reflect regional atrophy by nature, while functional or other pathophysiological process might occur before atrophy. PET has long been explored for AD diagnosis. And there are several modalities of PET using different radioligands including FDG-PET, AV45-PET, AV1451-PET, and so on.

In this study, I focus on using FDG and AV45 PET, which have been collected in a relatively large amount in ADNI. PET radioligand development is an active research area, and this study can be readily extended to other PET modalities in the future given abundant scans. In a similar way to interpret the final output from the model on structural MRI, the output from the model on PET will be a data-driven summary score of the PET pattern related to AD.

FDG-PET is one of the most abundant PET modalities and has a long history of application in various areas of clinical research (de Leon et al., 2001; Herholz et al., 2002; Susan M. Landau et al., 2011). AV45-PET has demonstrated correlation with the presence and density of β-amyloid in the brain at autopsy (Clark et al., 2011).

2D deep learning has previously been applied on FDG-PET (Ding et al., 2018) and showed good performance. In this study, I investigated using 3D convolutional neural network.

A.5.2 Methods

A.5.2.1 Data

The PET data was acquired from ADNI (Jagust et al., 2015). ADNI-1 includes FDG-PET and PIB-PET. ADNI-GO/ADNI-2 includes FDG-PET, AV45 (Florbetapir)-PET, and AV1451-PET in the late stage. ADNI-3 includes FDG-PET, AV45-PET, FBB (Florbetaben)-PET, and AV1451-PET. FDG-PET is the most abundant modality within ADNI. AV45-PET is the second most abundant modality. PIB-PET from ADNI-1 has just 228 scans from 103 subjects. As of 04/02/2019, FBB-PET has 239 scans from 236 subjects, AV1451-PET has 824 scans from 672 subjects, and are still actively acquiring, but the majority scans are from cognitively normal and MCI subjects.

In this study, I focused on FDG-PET from ADNI-1 and ADNI-GO/ADNI-2, AV45-PET from ADNI-GO/ADNI2. For FDG-PET, I included 1445 scans (484 AD, 961 CN) from 682 subjects (241 AD, 441 CN). For AV45-PET, I included 906 scans (180 AD, 726 CN) from 515 subjects (151 AD, 364 CN).

ADNI also provides prior-based composite scores of the PET measures derived from multiple ROIs. FDG composite score is the average FDG uptake of angular, temporal, and posterior cingulate (Susan M. Landau et al., 2011). AV45 composite score is the average AV45 SUVR of frontal, anterior cingulate, precuneus, and parietal cortex relative to the cerebellum (S. Landau & Jagust, 2015). These composite methods are based on the priors of AD pathology and prior imaging studies, instead of the data-driven way used in machine learning.

A.5.2.2 Analysis

Since different PET modalities have different numbers of scans, I performed two major analyses. In the first analysis, I performed individual analysis for each individual modality. This can take advantage of the data availability most effectively. In the second analysis, I gathered common scans for all modalities. In the present study, the modalities are FDG and AV45. But in the future, when the overlap among different modalities increase, it is possible and straightforward to include more modalities. Within this major analysis, I proposed two sub-analyses: the first one performing classification using one modality at a time, and the performance of different modalities can be compared; the second sub-analysis involves stacking the different modalities, to maximize the discriminative power. Practically, this poses more restriction on the acquisition side, and additionally need to ensure registration on the processing side. In ADNI, the AV45 scans were already co-registered to the baseline FDG scans, so it was possible to directly stack the two modalities. The registration quality was visually checked.

A.5.2.3 Modality Merge

In a deep learning model, the information from the two modalities can start to merge at different depth. The simplest method is to stack the two modalities at the input level, similar to the channels in an RGB image. However, despite the two modalities share the same spatial location, the inherent patterns in the individual modality could be different. Thus, I further explored the performance when merging the two modalities at different depth in the model. An illustration of the framework is shown in Figure A-18.

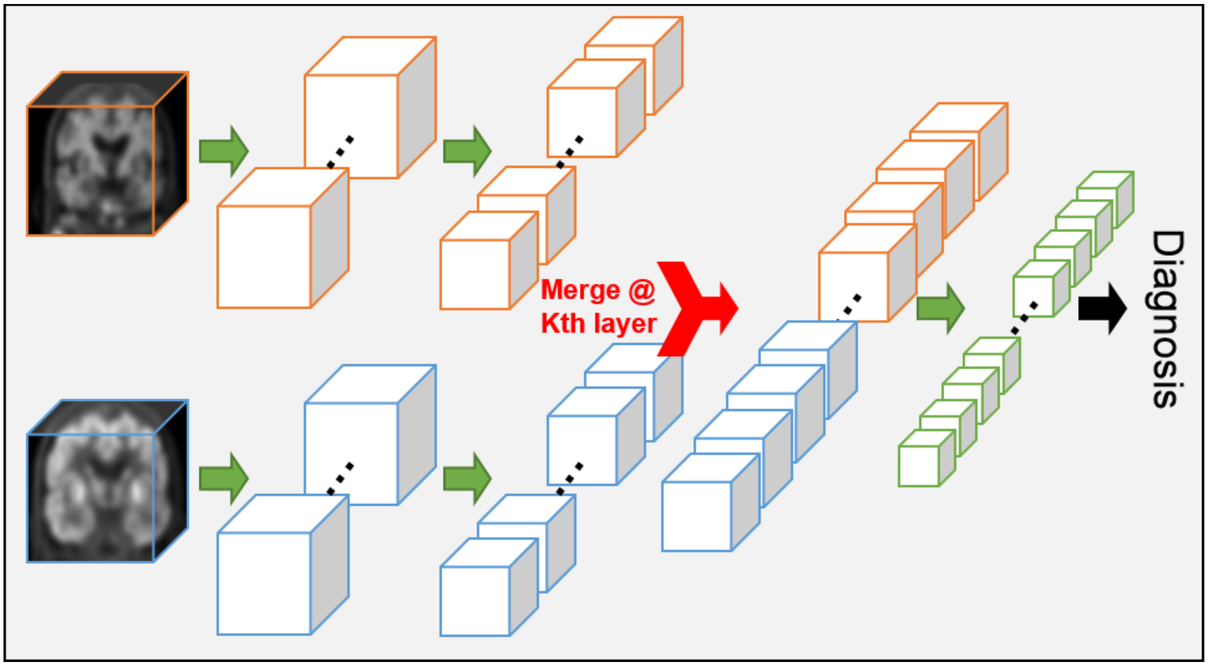


Figure A-18. Framework of merging features from multi-modal data at different depth of the network.

A.5.2.4 Normalization

Different from T1-weighted structural MRI, where the intensity of each individual pixel is a complicated reflection of the molecular composition of the underlying tissue and is not quantitative, the value of each pixel in PET represents the level of specific binding targets. The normalization step, a standard preprocessing step in most image processing tasks, remains to be explored. Thus, in this study, I present results using both normalized and raw input images.

A.5.2.5 DLPET scores

Similar to the generation of DLMRI score, I can generate deep learning derived PET (DLPET) scores for analyses. I analyzed the *DLFDG* scores from FDG-PET and *DLAV45* scores from AV45-PET in this study.

A.5.2.6 MCI progression prediction

I applied the model to the PET scans of baseline MCI subjects, and performed MCI progression prediction using ROC analysis and Cox proportional hazards regression survival analysis. The performance was compared with the PET composite measures and also DLMRI score derived from structural MRI. The table of demographical information is shown in Table A-3. The detailed analyses methods are similar and can be found in previous sections of DLMRI score analyses.

Table A-3. Demographical and summary information of the MCI progression dataset with both PET scans and PET composite measures available.

	MCI stable N = 154	MCI progression N = 92	Total N = 246
age	70.9 ± 7.0	72.9 ± 6.6	71.7 ± 6.9
sex M/F (%M)	88/66 (57.1)	51/41 (55.5)	139/107 (56.5)
APOE ε4 frequency (2/1/0)	11/44/99	17/49/26	28/94/126
Last visit FU year	4.43 ± 0.55	-	-
conversion year	-	1.94 ± 1.16	-

DLMRI score	0.376 ± 0.163	0.632 ± 0.191	0.472 ± 0.214
DLFDG score	0.109 ± 0.196	0.564 ± 0.362	0.279 ± 0.349
DLAV45 score	0.174 ± 0.272	0.639 ± 0.328	0.348 ± 0.370
FDG composite score	1.31 ± 0.11	1.16 ± 0.11	1.26 ± 0.13
AV45 composite score	1.15 ± 0.19	1.41 ± 0.21	1.25 ± 0.23

A.5.2.7 Association with the prior-based measures

I further evaluated the association between the DLPET scores and the prior-based composite scores through correlational analysis. I used all available MCI baseline measures, with N = 660 for FDG-PET, and N = 458 for AV45-PET.

A.5.2.8 Class activation map

The class activation maps were generated in a similar way as the one generated previously on structural MRI (Feng, Yang, et al., 2018). The average class activation map of AD cases was generated for interpretation. I also generated the average PET images of each PET modality to overlay with the class activation map to guide the localization in the class activation map.

A.5.3 Results

A.5.3.1 Individual modality classification performance

The test AUROC values and accuracies using individual PET modalities including FDG and AV45 are shown in Table A-4.

Table A-4. The classification performance using individual PET modalities.

	FDG	**AV45**
Normalized	.932/88.0%	.875/83.8%

Raw	**.910/92.0%**	.872/81.2%

A.5.3.2 Classification performance using common scans of FDG and AV45

The test AUROC values and accuracies using common FDG, AV45 scans, and stacking the FDG and AV45 at the input are shown in Table A-5.

Table A-5. Classification performance using common scans of FDG and AV45.

	FDG	**AV45**	**Stacking FDG, AV45**
Normalized	.876/93.3%	**.896/90.0%**	.889/83.3%
Raw	.891/91.7%	.893/86.7%	.821/80.0%

A.5.3.3 Classification performance merging FDG and AV45 features at different depth

The test AUROC values and accuracies stacking the FDG and AV45 feature maps at different depth are shown in Table A-6.

Table A-6. Classification performance stacking FDG and AV45 at different depth.

Merge level	**1**	**2**	**3**	**4**	**5**
Normalized	.928/91.7%	.872/91.7%	.885/91.7%	.882/90.0%	.854/90.0%
Raw	.923/90.0%	.901/91.7%	.896/93.3%	.892/91.7%	.879/90.0%

A.5.3.4 MCI progression prediction

I derived DLPET scores from two models with highest accuracies indicated as bold in Table A-4 and Table A-5. The ROC curves of MCI-progression and MCI-stable classification using composite PET scores, DLPET scores, and DLMRI score are show in Figure A-19, the AUROCs are shown in Table A-7. The Cox proportional hazards regression analyses statistics are shown in Table A-8. DLFDG scores show better time-to-conversion prediction performance. While DLFDG and DLAV45 scores show better ROC performance than DLMRI score with no covariates, the performance is inferior after controlling for covariates.

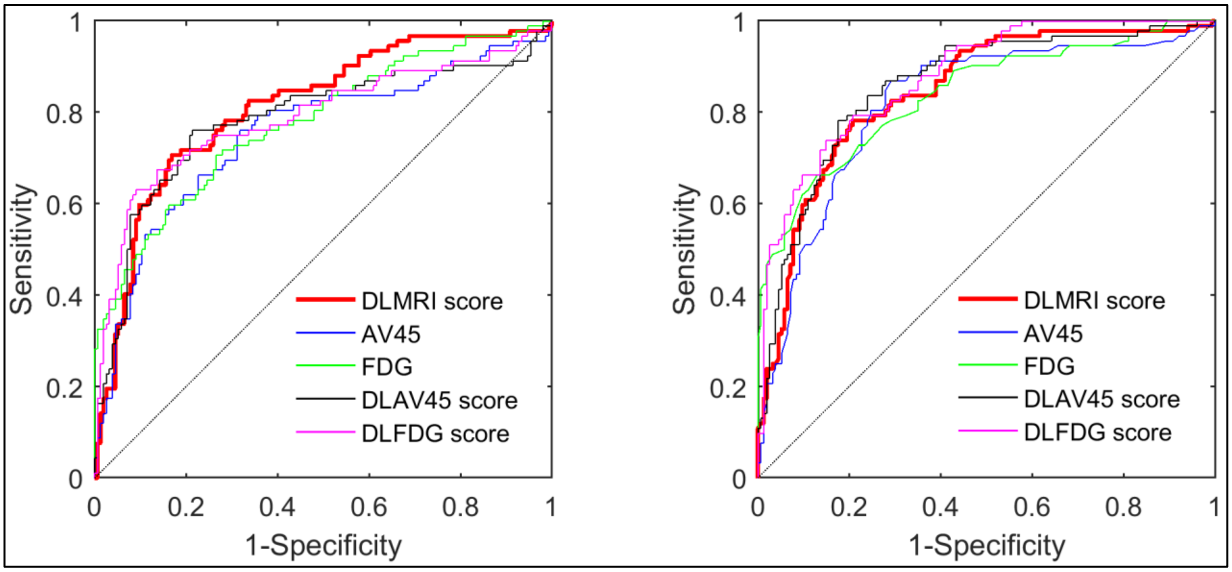


Figure A-19. ROC curves of different measures: (Left) Covariates age, gender, APOE ε4 frequency; (Right) No covariate.

Table A-7. AUROCs of the ROC analyses.

	No covariate	**Controlling for covariates**
AV45 composite	0.820	0.752
FDG composite	0.839	0.777
DLAV45 score	0.859	0.783
DLFDG score	**0.877**	0.789
DLMRI score	0.845	**0.810**

Table A-8. Cox proportional hazards regression survival analyses statistics.

	log HR	**SE**	**z**	**p-values**	**Chi-square**	**25% quantile**	**75% quantile**
AV45 composite	3.50	0.485	7.22	5.37E-13	49.6	1.03	1.43
FDG composite	-8.79	0.979	-8.98	2.76E-19	84.3	1.18	1.34
DLAV45 score	2.72	0.341	7.97	1.55E-15	68.7	0.017	0.742
DLFDG score	2.92	0.299	9.76	1.71E-22	95.3	0.013	0.555

DLMRI score	4.94	0.560	8.82	1.15E-18	81.8	0.297	0.623

A.5.3.5 Association with the prior-based composite scores

I showed the prior-based composite scores strongly correlated with the DLPET scores, suggesting the association and also discrepancy between the prior-based and data-driven summary scores of PET scans. The scatterplots between the two measures are shown in Figure A-20. As the distribution of the data is dense in the range of 0 – 0.05, I also zoomed in on this range, and performed correlation in this range. The correlation statistics are shown in Table A-9.

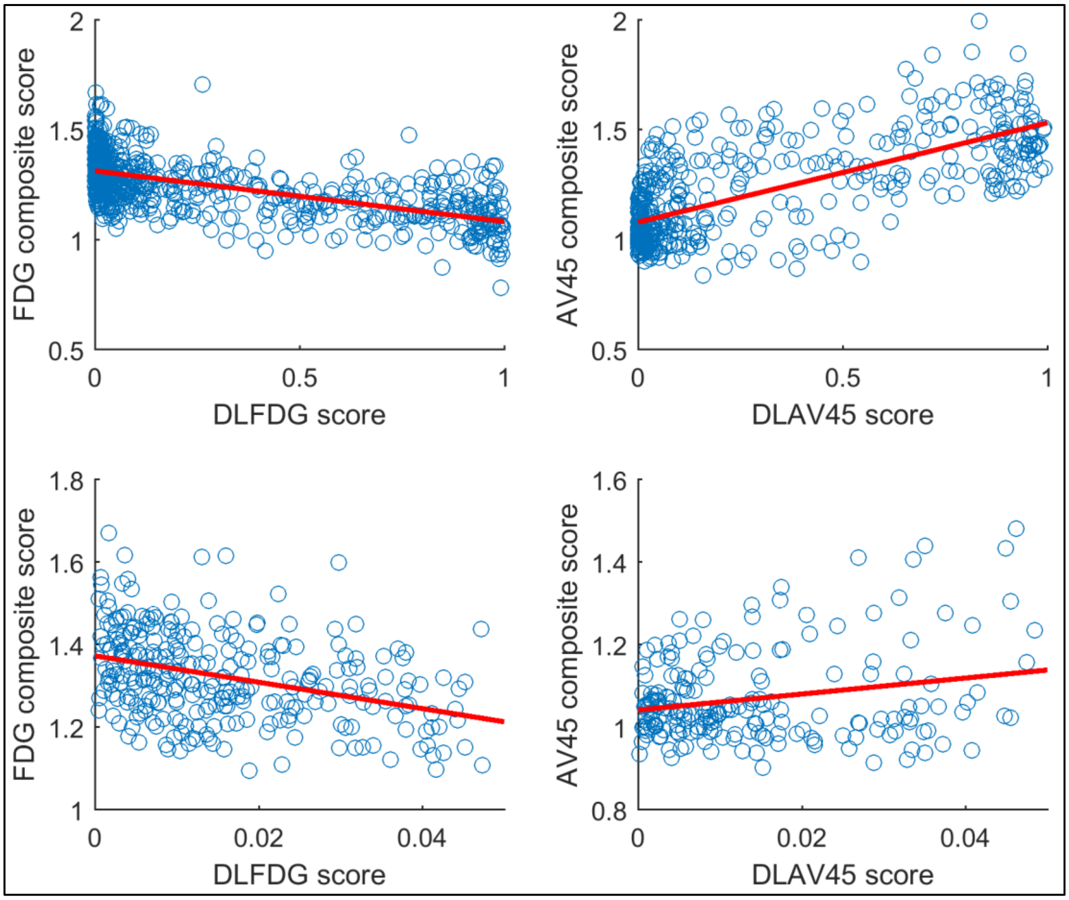


Figure A-20. Association between the deep learning derived PET summary scores and the prior-based composite scores. The red lines are the linear fits.

Table A-9. The correlation statistics between deep learning derived PET summary scores and the prior-based composite scores.

	corr-coef	**p-value**
FDG	-0.614	1.54E-69
DLFDG (0 - 0.05)	-0.368	1.41E-10
AV45	0.712	4.96E-72
DLAV45 (0 - 0.05)	0.230	1.10E-03

A.5.3.6 Class activation maps

The class activation map of the FDG-PET model is shown in Figure A-21. The predictive region localizes to medial occipitotemporal gyrus (lingual gyrus), inferior part of precuneus. The class activation map of the amyloid-PET model is shown in Figure A-22, one prominent region localizes to the right parietal regions, the other lesser prominent region localizes to left medial orbitofrontal region.

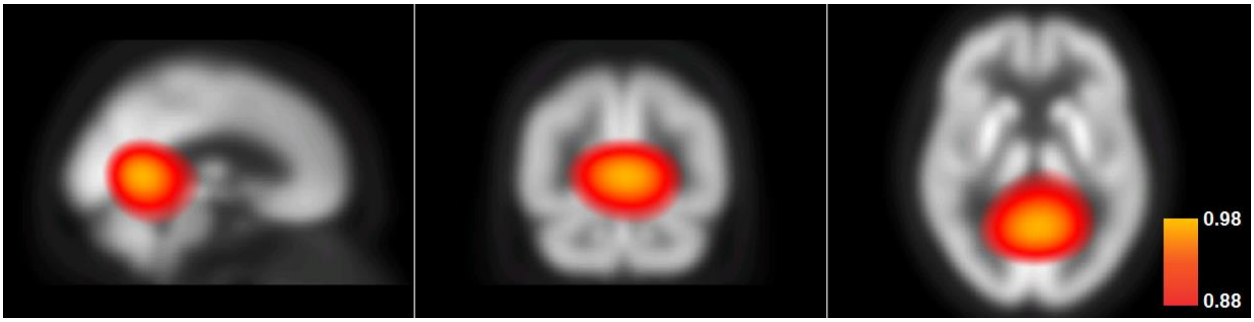


Figure A-21. The AD class activation map of the FDG classification model overlaid on the average FDG map from the population. The class activation map localizes to medial occipitotemporal gyrus and inferior part of precuneus.

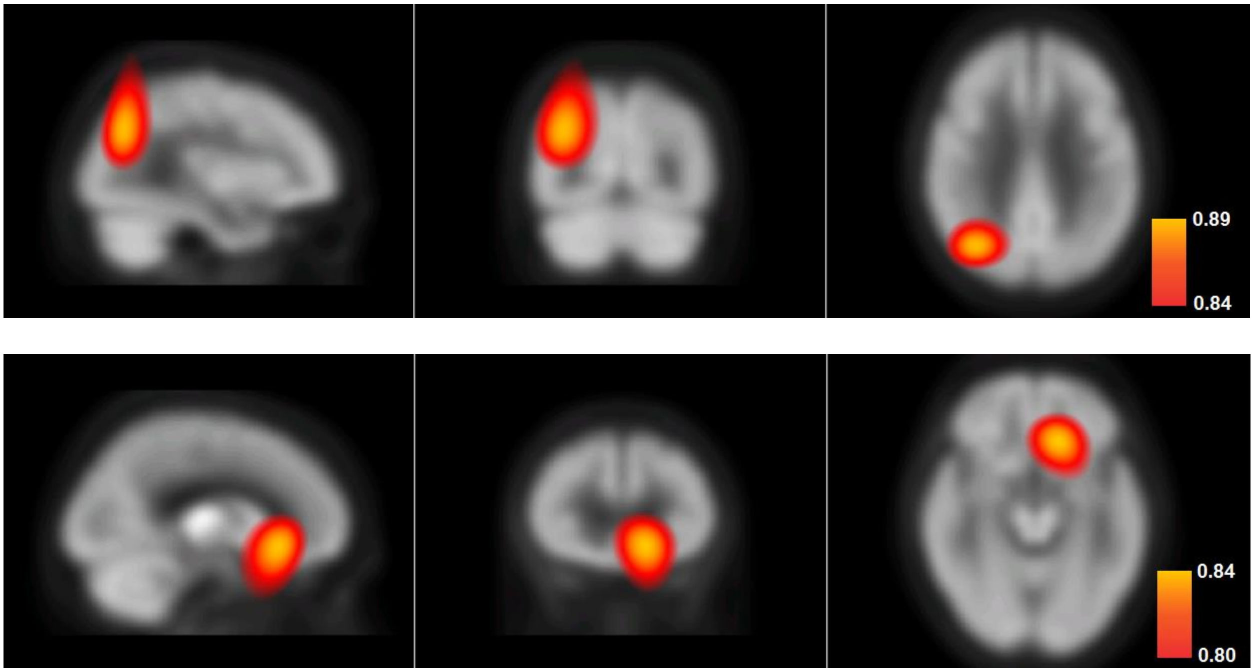


Figure A-22. The AD class activation map of the AV45 classification model overlaid on the average AV45 map from the population. There are two prominent blobs in the class activation map, the more prominent one localizes to right parietal lobe (top), the lesser prominent blob localize to left medial orbitofrontal lobe (bottom). Note the range of the color-maps are different in the two sub-figures to best visualize the individual blob.

A.5.4 Discussion

Individually, FDG achieves good performance, AV45 achieves moderate performance, possibly because of more abundant training data for FDG-PET. Using the same set of data does show comparable performance between AV45 and FDG. Simply stacking two modalities as two channels of data did not lead to better classification performance than using individual modality alone. Intuitively, it increases the input dimensionality without change to the number of samples. Whereas, merging the feature maps of the two modalities after the first convolutional block but not deeper layers improve the performance. One explanation could be that it is necessary to unify the latent representation of two modalities with some feature extraction operations before merging the information; however, it is beneficial to start the interaction between two modalities at lower representation.

The regions revealed through class activation map are consistent with previous studies and could potentially inform further neuroimaging studies. FDG hypometabolism in AD has been shown to localize to temporo-parietal regions (Hoffman et al., 2000). Amyloid accumulation revealed through AV45 Aβ has been shown to preferentially start in the precuneus, medial orbitofrontal, and posterior cingulate cortices (Palmqvist et al., 2017).

One limitation when interpreting the result is that the result is inevitably affected by structural information. This further complicates the attribution of the derived scores. However, on the other hand, this presents a natural way to merge different sources of information, which is important for biomarker or diagnostic tool development.

While the DLPET scores show better performance in ROC analysis without covariate than DLMRI score, the performance becomes inferior including covariates into the model. This could be potentially related to the sample composition.

One future work worth exploring is the explicit inclusion of interaction between two modalities into the model. Interaction is a term very commonly adopted in regression analysis, and can be explicitly included in the input. It could be potentially helpful for the classification if the interaction captures some pathophysiological signal, and it could also be helpful for interpretation purpose.

The other future work is to apply the model to other datasets, e.g. OASIS-3, which has both FDG and AV45 scans. But preprocessing standardization across studies requires further exploration.