

# Classification of Alzheimer's Patients Using Structural MRI Data

Yoon-Suk Han, Wyatt Hong

12 December 2013

## Introduction

Alzheimer's disease (AD) is a fatal neurodegenerative disease that currently affects more than 5 million people in the United States. It is the most common form of dementia whose symptoms include the degradation of personality, memory, perception, and cognitive skills. Patients with severe AD can no longer communicate, recognize their family, or perform the simplest daily activities. Due to population ageing, the number of AD patients in the US is expected to increase to 13 million by 2050, and AD related health care costs to \$500 billion.<sup>1</sup> Most patients go through a transitional stage called mild cognitive impairment (MCI) before they lapse into AD. MCI is characterized by a mild impairment in memory and executive function that is not severe enough to interfere with daily life, but which can be noticed by the individual and other people. Currently, clinical diagnosis by a physician is the only way for a patient to be diagnosed with either MCI or AD. However, the decreasing cost of genome sequencing, neuroimaging (structural and functional MRIs), and testing for molecular biomarkers in the plasma or the cerebral spinal fluid (CSF) has made a wealth of biomedical data available from each potential patient. Our study applied machine learning to the previously collected *structural MRI data* from the Alzheimer's Disease Neuroimaging Initiative (n = 1018) to diagnose a patient into one of the three categories: Normal, MCI, and AD. After applying feature selection to logistic regression models we were able to correctly classify patients with 76.34% accuracy for the binomial model (50% if randomly assigned) and 59.8% for the multinomial model (33.3% if randomly assigned). Our results illustrate the potential for machine learning models to complement and possibly replace the traditional model of clinical diagnosis made by doctors.

## Methods

### *Data Collection*

A major source of public data for AD study is the Alzheimer's Disease Neuroimaging Initiative (ADNI). ADNI is a longitudinal, multi-center study that collects data from AD patients all over the United States. The dataset, which comprises biomarkers from more than 1,000 subjects, includes personal history, clinical/cognitive assessments, genome sequence, magnetic resonance imaging (MRI), and positron emission tomography (PET) scans. We specifically worked with the T1-weighted structural MRI scans (<http://adni.loni.ucla.edu/>) of 292 healthy older controls, 549 mild cognitive impairment patients, and 176 AD patients (total of 1018 subjects). All data was collected according to the Good Clinical Practice guidelines, the Declaration of Helsinki, and U.S. 21 CFR Part 50—Protection of Human Subjects, and Part 56—Institutional Review Boards.

### *Data Processing*

For each subject, we extracted the structural T1 weighted MRI images acquired at inclusion of the study. To correct for the differences in brain size and shape across the subject population, we then spatially normalized the brain images to the MNI space by using FSL software (<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>). More precisely, we first used a linear co-registration implemented in FSL's FLIRT to compute a good starting point for the more accurate non-linear registration implemented in FSL's FNIRT (Figure 1).

### *Feature Extraction*

Once all images had been spatially normalized, we used a region of interest (ROI) approach and computed the average gray matter density (indirect measure of the number of neurons/volume) in each of the 115 distinct brain regions as defined by the Automatic Anatomical Labeling Atlas (AAL) (Figure 1).<sup>2</sup> These 115 values, labeled ROI\_1 through ROI\_115, were our primary features.

clinical diagnosis (Normal, MCI, or AD) given by a physician.

### Machine Learning Algorithm

We constructed two models for our study (Figure 2). The goal of the *binomial model* was to distinguish between normal and diseased (either MCI or AD) patients while that of the *multinomial model* was to distinguish among all three disease categories (Normal, MCI, and AD). We implemented logistic regression for both models using *scikit-learn Python Machine Learning Library*.<sup>3</sup> In the case of the multinomial model, we used a one-vs-all scheme to train the data.

### Feature Selection

We used forward search to find the subset of features that performs best for each of our two models. With the reduced set of features, we saw about 1~2% improvement in each model's performance, measured by its accuracy.

### Calculating Accuracy

To evaluate the performance of our models, we used a 10-fold cross validation on the entire data set. All accuracy percentiles noted in this paper refer to the results of cross-validation.

## Results

### Binomial Model

After feature selection, we achieved 76.34% accuracy for the binomial classification between Normal and Diseased (either MCI or AD) patients using logistic regression. The feature set used in this model includes ROI\_38, ROI\_44, ROI\_13, ROI\_65, APOE4, PROTOCOL, ROI\_108, ROI\_1, and ROI\_2 in the descending order of the absolute value of feature weight.

### Multinomial Model

After feature selection, we achieved 59.8% accuracy for the multinomial classification among Normal, MCI and AD patients using logistic regression. The feature set used in this model includes PROTOCOL, APOE4, ROI\_10, ROI\_13, ROI\_38, ROI\_41, ROI\_45, ROI\_51, ROI\_54, ROI\_71, ROI\_93 (not listed in order). Since the multinomial model uses a one-vs-all scheme, it computes three decision functions to classify each of our classes versus the rest. Feature weight rankings for the three decision functions were as follows:

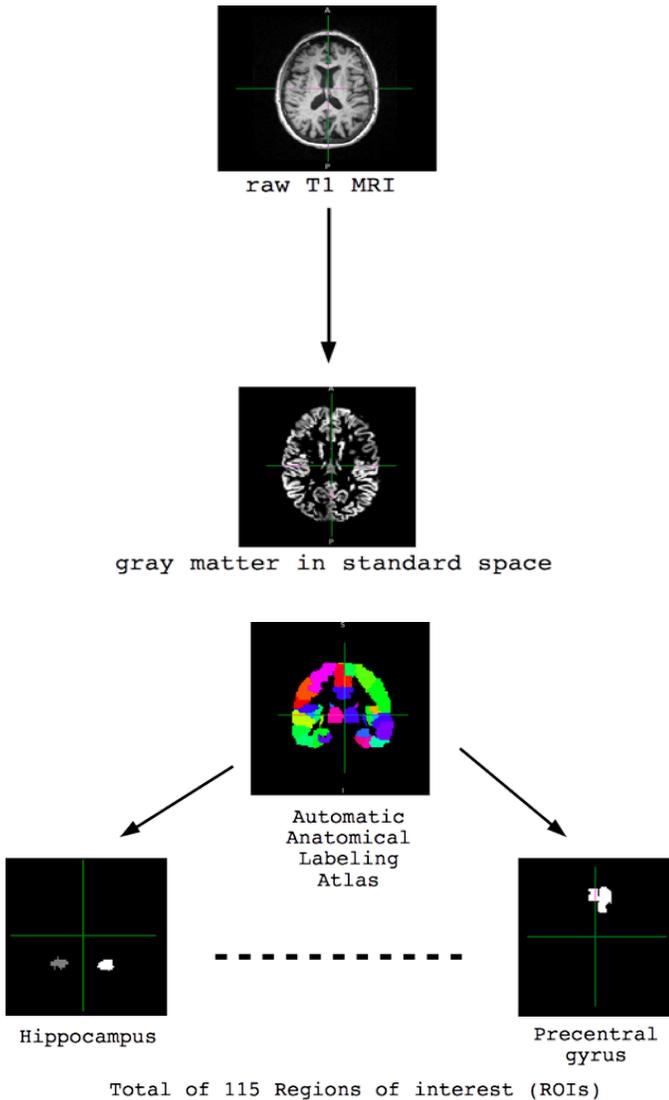


Figure 1. MRI data processing pipeline. Top: Raw T1 structural images from 1018 subjects were spatially normalized into standard space then filtered for voxels corresponding to gray matter. Bottom: The extraction of average gray matter in each of the 115 brain regions defined by the AAL (each color represents a distinct region). The hippocampus and the precentral gyrus are shown as examples.

### Additional Features

We included four additional features to improve our model: 1) *Age* – Age is the biggest risk factor for AD (However, there is no significant between group difference in age in the ADNI cohorts) 2) *APOE allele status* – the e4 allele of the APOE (apolipoprotein E) gene is the most significant genetic risk-factor for late-onset AD 3) *Sex* – the risk of AD has been shown to be modulated by gender 4) *Protocol* – the phase of the study was included as a feature to correct for the differing patient population between the two phases of ADNI (ADNI1 vs. ADNI2). Thus, we arrived at a total of  $115 + 4 = 119$  features for our 1018 subjects ( $n = 119$ ,  $m = 1018$ ). The label for each subject was his or her

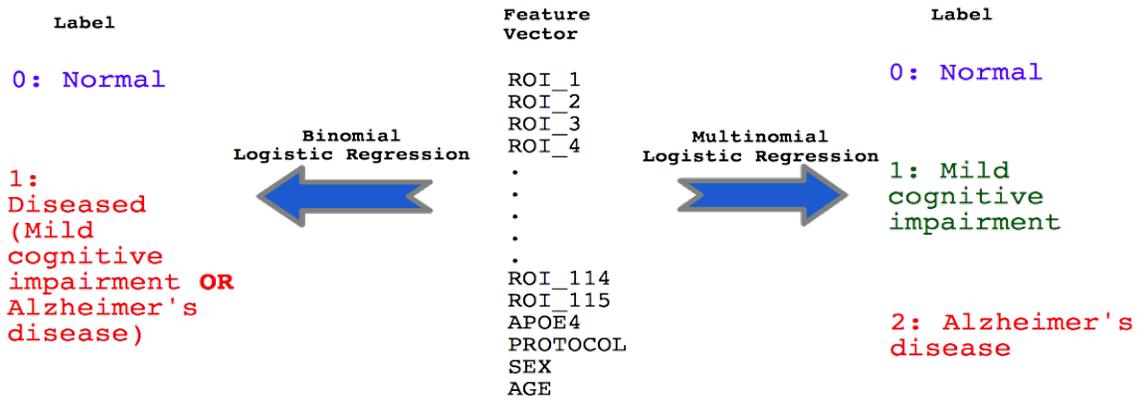


Figure 2. Overview of the classification task

## Discussion

Using structural MRI data, we were able to correctly classify patients 76.34% for the binomial model (50% if randomly assigned) and 59.8% for the multinomial model (33.3% if randomly assigned). A two-dimensional visualization of both models using PCA shows that the top features are able to separate the groups, especially in the multinomial model (Figure 3). Our work shows that machine learning can be applied to neuroimaging data to successfully mimic a doctor's diagnosis. Given that a structural MRI scan takes 15-30 minutes per patient, our method presents a faster, more objective method of diagnosis.

Observation of feature weights reveals the presence of features that are consistently significant across all models. These features correspond to the following brain regions as defined by the Automatic Anatomical Labeling Atlas:

- ROI\_38 - Right hippocampus*
- ROI\_13 - Inferior frontal gyrus*
- ROI\_44 - Precuneus cortex*
- ROI\_45 - Precuneus cortex*
- ROI\_41 - Left amygdala*

Not surprisingly, previous molecular and clinical studies have shown many of these regions to be associated with AD and other neurological diseases that cause cognitive deficiency. The hippocampus serves as the seat for long-term memory and is the first region to degenerate in AD patients. Hippocampal atrophy has been associated with AD since the beginning of molecular studies in humans and mice.<sup>4</sup> The inferior frontal gyrus has been associated with AD, major depression, and deficits in attention tasks.<sup>5</sup> The precuneus cortex is heavily involved in memory recall and its atrophy has been associated with

Normal vs Rest

ROI\_38, ROI\_41, ROI\_13, APOE4, ROI\_71, ROI\_51, ROI\_45, ROI\_10, ROI\_93, PROTOCOL, ROI\_54

MCI vs Rest

ROI\_38, ROI\_13, ROI\_45, PROTOCOL, ROI\_71, ROI\_51, ROI\_10, APOE4, ROI\_41, ROI\_54, ROI\_93

AD vs Rest

ROI\_38, ROI\_41, ROI\_71, ROI\_51, PROTOCOL, ROI\_45, ROI\_93, APOE4, ROI\_54, ROI\_13, ROI\_10

## SVM

To compare the performance of logistic regression with other algorithms, we also tried to fit our training data using SVM with regularization parameter ( $C = 1$ ). Again, we constructed two models, binomial and multinomial. The multinomial classification was implemented via a one-vs-one scheme. We tried both a 3rd-degree polynomial kernel and a linear kernel to see whether there was any difference in performance between the two kernels. The results are shown in Table 1.

	Multinomial	Binomial
Polynomial kernel	59.51%	76.14%
Linear kernel	60.89%	76.53%

Table 1. Accuracy of SVM models

The linear kernel performs slightly better than the polynomial kernel. As expected, the polynomial kernel was prone to over-fitting (its test error was much greater than the training error regardless of the size of the training set). Overall, we did not see a significant difference in performance between SVM and logistic regression, showing us that we have maximized the performances of our models.

early-onset AD.<sup>6</sup> The amygdala is involved in fear-response and emotional memory and its atrophy has been associated with AD as well.<sup>7</sup> The identification of these previously known regions with our models further gives credence to the applicability of machine learning in the field of medical diagnosis.

learning models may be able to outperform and replace the traditional model of clinical diagnosis.

### Acknowledgements

Andre Altmann ([altmann@stanford.edu](mailto:altmann@stanford.edu))  
 Post-doctoral fellow, FIND Lab, Stanford Medical School,  
 Stanford, CA.

### References

1. Bateman, R. J. *et al.* Clinical and Biomarker Changes in Dominantly Inherited Alzheimer's Disease. *N. Engl. J. Med.* **367**, 795–804 (2012).
2. Tzourio-Mazoyer, N. *et al.* Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* **15**, 273–289 (2002).
3. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J Mach Learn Res* **12**, 2825–2830 (2011).
4. Hyman, B. T., Hoesen, G. V., Damasio, A. R. & Barnes, C. L. Alzheimer's disease: cell-specific pathology isolates the hippocampal formation. *Science* **225**, 1168–1170 (1984).
5. Hirono, N. *et al.* Frontal lobe hypometabolism and depression in Alzheimer's disease. *Neurology* **50**, 380–383 (1998).
6. Karas, G. *et al.* Precuneus atrophy in early-onset Alzheimer's disease: a morphometric structural MRI study. *Neuroradiology* **49**, 967–976 (2007).
7. Poulin, S. P. *et al.* Amygdala atrophy is prominent in early Alzheimer's disease and relates to symptom severity. *Psychiatry Res.* **194**, 7–13 (2011).

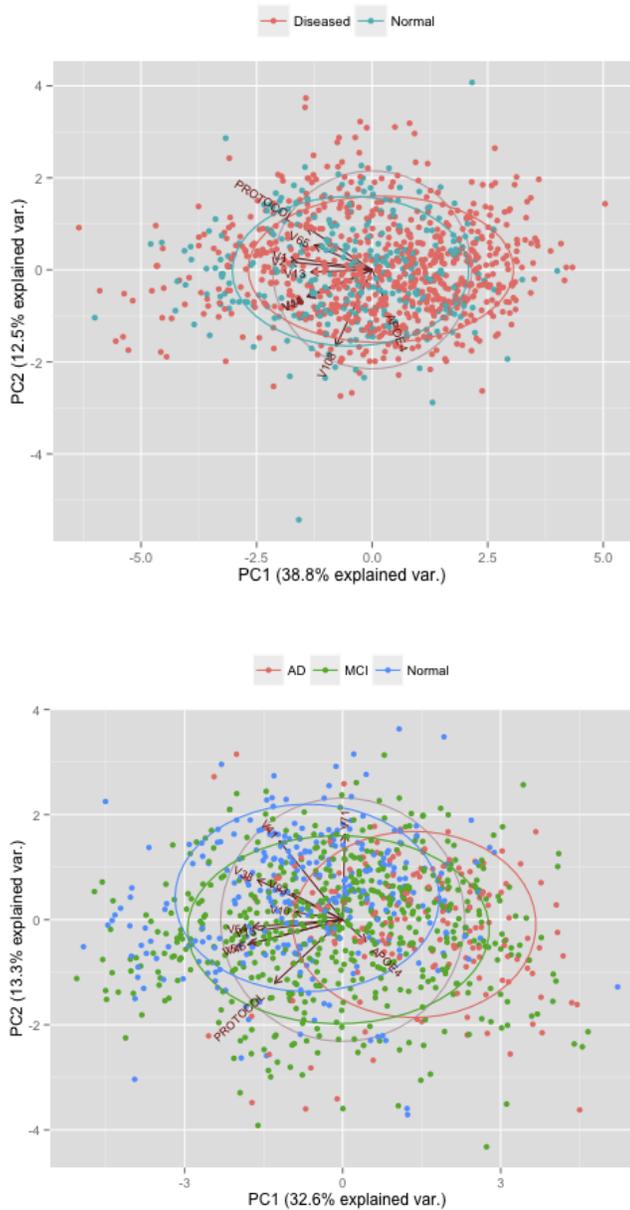


Figure 3. Plots created using the ggbiplot package in R where the first principal axis is the horizontal axis and the second principal axis is the vertical. Top: Visualization of binomial data (Blue = Normal, Red = Diseased). Bottom: Visualization of multinomial data (Blue = Normal, Green = MCI, Red = AD).

Combining structural MRI data with additional neuroimaging data, including functional MRI (fMRI) and positive emission tomography (PET), may further increase the performance of the model until machine