Predicting Dementia Diagnosis with Neuroimaging Data

Boya Jiang,Chen Chen,Tiankai Xie,Yuan Zhong

12/15/2021

Introduction

Alzheimer's is a dementia degenerative disease starting with mild memory impairment in the early stages and progressing to a complete loss of the mental and physical faculties. Definitive Alzheimer's Disease (AD) diagnosis relies on a magnetic resonance image (MRI) study. Brain MRI scans are detailed three-dimensional anatomical images, and changes in the hippocampus, frontal and parietal regions are evidential markers in the progress of AD. The ability to diagnose and classify AD at an early stage allows clinicians to make more knowledgeable decisions regarding clinical interventions. In this project, we apply multiple representative methods to predict Alzheimer's status.

Data and Preprocessing

OASIS (Open Access Series of Imaging Studies) is a well-known initiative that is publicly available for study and analysis. The present MRI dataset, OASIS-I (presented in 2007), is a cross-sectional collection of data for 416 participants aged 18-96 yrs, 316 non-demented and 100 at various stages of AD. Subjects were characterized by the Clinical Dementia Rating (CDR) scale from cognitive normal (CDR = 0), very-mild dementia (CDR = 0.5) to mild dementia (CDR = 1). The data set also contains the following demographics information: male/female, age, education (Educ), and socioeconomic status (SES). Data set of MRI scans of axial plane contains 176 slices/images of 176 x 208 pixel size.

Since AD is more prevalent among older adults, we selected individuals between 60 and 96 years old. The CDR was dichotomized to 0 for cognitive normal (CN), and 1 for any level of dementia. Our final samples composed of 198 individuals, which were randomly split into two groups in 80:20 ratios as training and testing sets, respectively. The training and testing sets had balanced CDR distribution. Table 1 shows a summary of demographic variables collected for CN and dementia groups. [**Description of SES missing data handling**]. Figure 1 shows a comparison of MRI scan between a CN patient and a dementia patient at the same slice.

Methods and Results

In order to classify dementia from CN, we applied the following four methods: penalized logistic regression (ridge and lasso), random forest, support vector machine (linear and radial kernel), and convolutional neural network (CNN).

a) Logistic Regression
Logistic regression (LR) was perferred over linear discriminant analysis (LDA) because it does not require the independent variables to satisfy the assumptions of linearity, normal distribution, or equal variance. LR also provides a deterministic model yielding weighting factors for each contributing variable, while avoiding overfitting the data.

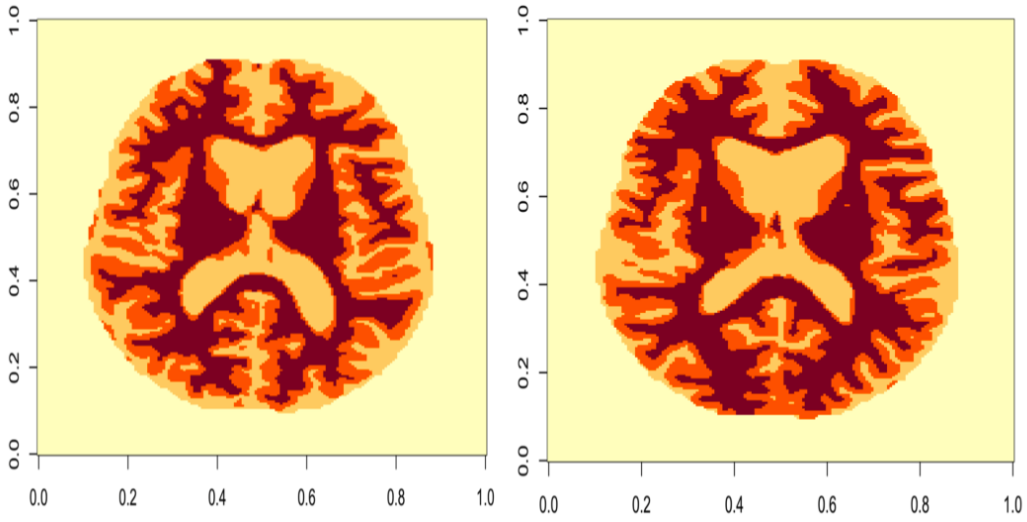

Figure 1: MRI scan of CN (left) and dementia (right) patients

In order to prepare data for logistic regression, we selected the middle slice for each patient, creating an array of dimensions 176x208x176. Next, to improve the speed of model analysis, data was further cleaned by removing axes containing all zero values. Parameter tuning was performed through 10-fold cross-validation analysis via `cv.glmnet`. Both ridge and lasso methods were used for regularization. On that basis, we further incorporated demographic variables with neuroimaging data and fit ridge logistic regression.

b) Random Forest

c) SVM

b) CNN

Conclusion and Discussion

Deliverable and Code Repository

Shiny App

Shiny Application We built our Shiny application using the CNN model(?) as our classifier embedded behind the dashboard. This means that the dashboard can take patients MRI data and some demographic and clinical input to predicts of Dementia diagnosis using CNN model. The Shiny application can be viewed by running this command, shiny::runGitHub('625OASIS','y1zhong', subdir = 'shiny'), on the R console. Packages like Shiny and oro.nifti must be installed before this app can be run.

GitHub:

Google Colab?

Contributions

References