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 [1]. 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) [2], 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 sample composed of 198 individuals, which were randomly split into two groups in 80:20 ratios as training and testing sets, respectively. For our variable of interest, there does not exist data formatting problem, but there is a missing data problem in SES, with 18 missing records. By using predictive mean matching as imputation method from mice package, missing data problem was solved. Table 1 shows a summary of demographic variables collected for CN and dementia groups. The training and testing sets had balanced CDR distribution. Figure 1 shows a comparison of MRI scan between a CN patient and a dementia patient at the same slice.

Methods

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

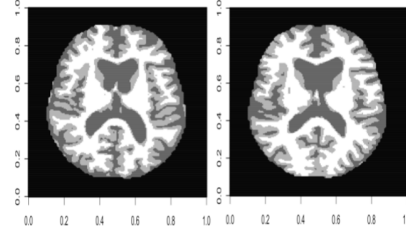| Characteristic	Overall, N = 198	Cognitive Normal, N = 98	Dementia, N = 100
M.F			
F	131 (66%)	72 (73%)	59 (59%)
M	67 (34%)	26 (27%)	41 (41%)
Age	76 (71, 82)	74 (69, 83)	77 (72, 81)
Educ			
1	23 (12%)	8 (8.2%)	15 (15%)
2	60 (30%)	25 (26%)	35 (35%)
3	40 (20%)	23 (23%)	17 (17%)
4	36 (18%)	17 (17%)	19 (19%)
5	39 (20%)	25 (26%)	14 (14%)
SES			
1	37 (19%)	20 (20%)	17 (17%)
2	52 (26%)	31 (32%)	21 (21%)
3	49 (25%)	23 (23%)	26 (26%)
4	57 (29%)	23 (23%)	34 (34%)
5	3 (1.5%)	1 (1.0%)	2 (2.0%)

Table 1: Summary of Demographics

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

Data Preperation In order to prepare data for all 3D methods (a to d), we selected the middle slice for each patient, creating an array of dimensions 176 x 208 x 198. Next, to improve the speed of model training, data was further cleaned by removing axes containing all zero values. For logistic regression, parameter tuning was performed through 10-fold cross-validation analysis via `cv.glmnet()`. For random forest, SVM, and native Bayes, principal component analysis (PCA) was performed for feature extraction prior to model fitting. In addition, parameter tuning was performed through 10-fold cross-validation analysis via `caret::trainControl()`, which controls the parameters for train data. For methods a) to d), models were fit on neuroimaging data with and without demographic variables.

a) Logistic Regression (LR)
Logistic regression was preferred 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. Both ridge and lasso methods were used for regularization.

b) Random Forest (RF)
Random forest is an ensemble learning method. For classification tasks, its output is the class selected by most trees. It doesn't overfit with more features and its efficiency is particularly notable in large data sets. The random forest method was called from `randomForest`.

c) Support Vector Machine (SVM)
Support vector machine is a supervised learning method. It is effective in high dimensional spaces and in cases where number of dimensions is greater than the number of samples. It is also memory efficient since it uses a subset of training points in the decision function, known as support vectors. Both linear and radial kernel were used for regularization. The SVM with linear kernel was called from `e1071` and the SVM with radial kernel was called from `kernlab`.

d) Naive Bayes
Naive Bayes is a classifier based on applying Bayes' theorem with strong independence assumptions between the features. It couples simple Bayesian network models with kernel density estimation, and can achieve higher accuracy levels. Unlike ther types of classifiers wihch use expensive iterative approximation, its maximum-likelihood training can be done by evaluating a closed-form expression, which only takes linear time. The Naive Bayes was called from `klaR`.

e) Convolutional Neural Network (CNN)
To avoid information loss from 3-d images to 2-d matrix in above methods , we tried to employ convolutional neural network, which is a popular deep learning technique in medical imaging analysis. VGG is a deep learning architecture shown to be able to reach top accuracy in image classification tasks. It utilizes 3-d filters and fully connected layers that usually need hundreds of millions of parameters for a successfully trained model. In this project, we mimicked VGG16 with 16 layers, 1 millions parameters. We used Keras library in Python script and GPU in Google Colab to efficiently train our model. However, we did not find a meaningful train feature that has the discernibility to classify test data into correct diagnosis results.

With several modified architectures, all the accuracies were around or below 50%. We did not report the evaluation here and left it for future exploration.

Deliverable

Shiny App

Shiny Application We built our Shiny application using the ridge LR as our classifier embedded behind the dashboard. This means that the dashboard can take patients' MRI data coupled with demographic and clinical input to predict Dementia status. The Shiny application can be viewed by accessing through 'shinyapps.io', https://j3z4te-tiankai-xie.shinyapps.io/shiny/, and a workable input called myMRI.zip could be found under shiny folder. Installation of Shiny and oro.nifti are required to run this app.

Challenges

Given the high dimensional structure of our data, we ran into several problems while attempting to train models on local environment. Specifically, the RF and SVM methods failed due to stack overflow error. We were able to resolve this problem by training the models on the Biostatistics Cluster. In addition, as CNN was a computationally demanding method, its initial runtime on the Biostatistics Cluster took approximately 40 hours. In order to improve the performance, CNN was trained on Google Colab with Graphics Processing Unit (GPU) accelerator. As a result, the runtime was reduced to approximately 30 minutes.

Evaluation

To evaluate the performance of each model, we compared test sensitivity, specificity, and accuracy (Table 2). When only image data was used to train the model, the best classification model was Ridge regression with the highest accuracy (0.79), sensitivity (0.75), and specificity (0.83). Figure 2 presented the ROC curve for each model which also indicated that Ridge regression was the best model to predict Alzheimer dementia. Lasso regression also performed well with a relative high accuracy that correctly classified 68% of the MRI. However, RF and LDA had a higher sensitivity than Lasso regression (0.65 and 0.63 respectively). The purpose of our analysis was to correctly predict AD, so sensitivity was more important than accuracy which indicated the percentage of true positive among test cases. We further included demographic information in our models, but the performance of prediction was not improved and some RF even had 28% decrease in sensitivity.

	Image only						Image+CS				
	Lasso	Ridge	LDA	SVM-L	RF	NB	Ridge	LDA	SVM-L	RF	NB
Sensitivity	0.6	0.75	0.5	0.5	0.5	0.6	0.75	0.6	0.5	0.5	0.6
Specificity	0.78	0.83	0.33	0.28	0.61	0.5	0.83	0.67	0.28	0.72	0.5
Accuracy	0.68	0.79	0.42	0.39	0.55	0.55	0.79	0.63	0.39	0.61	0.55

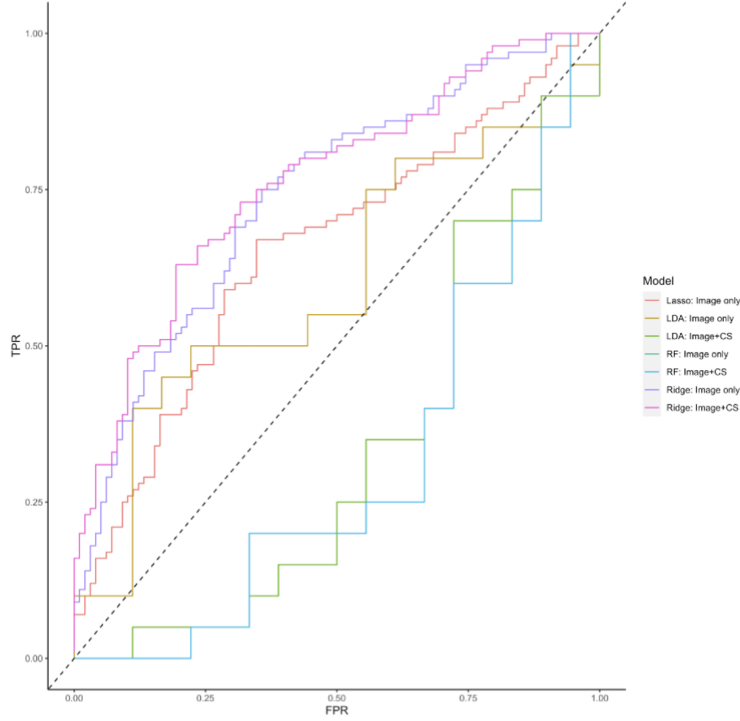Table 2. Classification results of sensitivity, specificity, and accuracy

Figure 2: ROC curves of selected models

Discussion

In conclusion, we compared the prediction of multiple classification models, and the Ridge regression outperformed others in terms of accuracy, sensitivity, and specificity. Comparing to literature, our analyses had several important strengths. Some literature may have outstanding accuracy that is more than 95% when comparing MRI of Alzheimer's dementia with healthy control in young adults[4]. However, young adults were not at risk of developing AD and using their MRI as control would introduce bias and over predict the result. Our analysis only included older adults which would be more practical to distinguish AD and healthy aging brain. Secondly, our outcome AD was balanced among cases and controls which would increase our accuracy. However, accuracy of AD prediction was different by the stages of AD. Since there were no significant symptoms at early AD stage when the degeneration of the brain had begun, it was difficult to diagnosis AD and more likely to discover AD at later stages. Therefore, accuracy would be differentiated by the level of comparison and distinguishing MCI from healthy aging would be the most difficult. Past classification models showed much lower accuracy between 40% to 60%. In our analysis, we combined MCI, AD, and severe AD as one group which made prediction harder with lower accuracy. Our CNN method was the most challenging model as it not only required large sample size, but also advanced computing environment such as cluster computing or GPU computing to incorporate the high-dimensional structure. Many research teams already recruited extraordinary engineers and computer science expertise to improve the algorithm, but with less focus on important biomarkers of AD such as APOE and CSF and other important clinical information. Future research should consider including more biostatistics, neurology, and epidemiology field to create multi-disciplinary research.

Contributions

Boya Jiang: LR, Data Visualzation
Chen Chen: Data Visualization, Literature Review, Evaluation
Tiankai Xie: SVM, RF, native Bayes, Shiny App
Yuan Zhong: CNN, Data Preprocessing
All members contributed in drafting the report, and each group member contributes to all other tasks as well

References

1. Alzheimer's Association. 2016 Alzheimer's disease facts and figures. Alzheimer's Dementia. 2016, 12, 459–509.

2. Open Access Series of Imaging Studies (OASIS): Cross-Sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults. Marcus, DS, Wang, TH, Parker, J, Csernansky, JG, Morris, JC, Buckner, RL. Journal of Cognitive Neuroscience, 19, 1498-1507. doi: 10.1162/jocn.2007.19.9.1498.

3. 3D Convolutional Neural Networks for Diagnosis of Alzheimer's Disease via Structural MRI. E. Yagis, L. Citi, S. Diciotti, C. Marzi, S. Workalemahu Atnafu and A. G. Seco De Herrera, 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), 2020, pp. 65-70, doi: 10.1109/CBMS49503.2020.00020.

4. Pellegrini, Enrico, et al. "Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: a systematic review." Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 10 (2018): 519-535.