The Use of MRI Data to Predict Dementia

Group 30

James Sinclair, ***867, jsinclair32@gatech.edu

04/14/2023

ISYE 7406

Abstract

Dementia is a brain disease that affects cognitive function, but it is difficult to diagnose. The OASIS project provides MRI and patient related data with the hope of furthering the understanding of the human mind. This paper aims to find a model that can best predict a patient's Clinical Dementia Rating. Eight different modeling methods were used with the data, split into training and testing sets, to develop models and then test their performance. It was found that with this dataset, the PLS modeling method produced a model that had the best performance with prediction accuracy. There is still much more data to be collected from patients and testing to be done, but definite progress has been made to solve the issue of diagnosing a patient.

(i) Introduction

The Open Access Series of Imaging Studies (OASIS) is a project that provides a collection of magnetic resonance imaging (MRI) and related data to the scientific community to further the study of basic and clinical neuroscience. The goal of this paper is to answer whether variables from the OASIS dataset can be used to accurately predict a patient's Clinical Dementia Rating (CDR)? In order to try and answer this question, this paper combines data sets from the OASIS website, explores the resulting dataset for insights, develops and tests 8 different models, and evaluates the performance of each model to determine which would be best for predicting a patient's CDR. Finally, a reflection on further research analysis and recommendations for the OASIS team are provided along with other concluding thoughts.

(ii) Problem Statement & Exploratory Data Analysis

Dementia is a type of disease of the brain that is often characterized by impairment of memory, progressive loss of intellectual thought and abstract thinking, and sometimes personality change. Currently, there is no definite test to determine if a particular patient has dementia. Doctors can determine generally if a person has dementia, but this is often based on a look at their medical history, lab tests, and reports of the patient's behavior (which requires someone else to report and explain)[1]. In general, it is a difficult process to diagnose dementia and even then, it can be difficult to determine the exact level of dementia, with cases of misdiagnoses being common[2]. An accurate model that can predict a patient's CDR would not only save time and resources for hospitals, it will also result in faster and better care of the patient that needs it.

There are two datasets provided by OASIS. The first dataset consists of data taken from cross-sectional images of 416 subjects aged 18 to 96. The older adults in this dataset are a mix of both "nondemented" or free of disease, and "demented" or having Alzheimer's disease. Meanwhile, the second dataset consists of data taken from longitudinal images of 150 subjects aged 60 to 96. MRIs are often taken in two planes, the axial plane (cross-sectional images) and the sagittal plane (longitudinal images) for diagnostic purposes. By taking MRIs in two planes, doctors can get a more comprehensive view of the internal structure being imaged. Additionally,

doctors can more easily detect subtle abnormalities that may not be visible in a single plane. This set also has a mix of both “nondemented” and “demented” subjects. Both of the first and second datasets share a few variables, however not all of the same variables were taken for each set. Below is a brief summary of each variable that the two sets have in common:

- CDR: Clinical Dementia Rating, scale 0-3, 3 characterizing the strongest onset of dementia
- M.F: patient is either male or female
- Age: age of patient
- Educ: years of education of patient, 1-23
- SES: Socioeconomic status, scale 1-5 (based on income)
- MMSE: Mini Mental State Examination, 11-questions that test five areas of cognitive function: orientation, registration, attention and calculation, recall, and language, scored 0-30
- eTIV: Estimated total intracranial volume, estimated from MRI images
- nWBV: Normalized whole brain volume, allows for averaging data across subjects by establishing spatial correspondence between brains
- ASF: Atlas scaling factor, factor used for normalizing WBV
- Section: indicates whether the eTIV comes from a cross-sectional or longitudinal MRI image

Upon combining the two sets, the patients with incomplete data were removed, resulting in a total of 570 rows of patients with the above variables.

Initial exploration into the data was conducted by creating and evaluating a correlogram. In order to visualize any potential relationships between *CDR* and any other variables. It is pictured below:

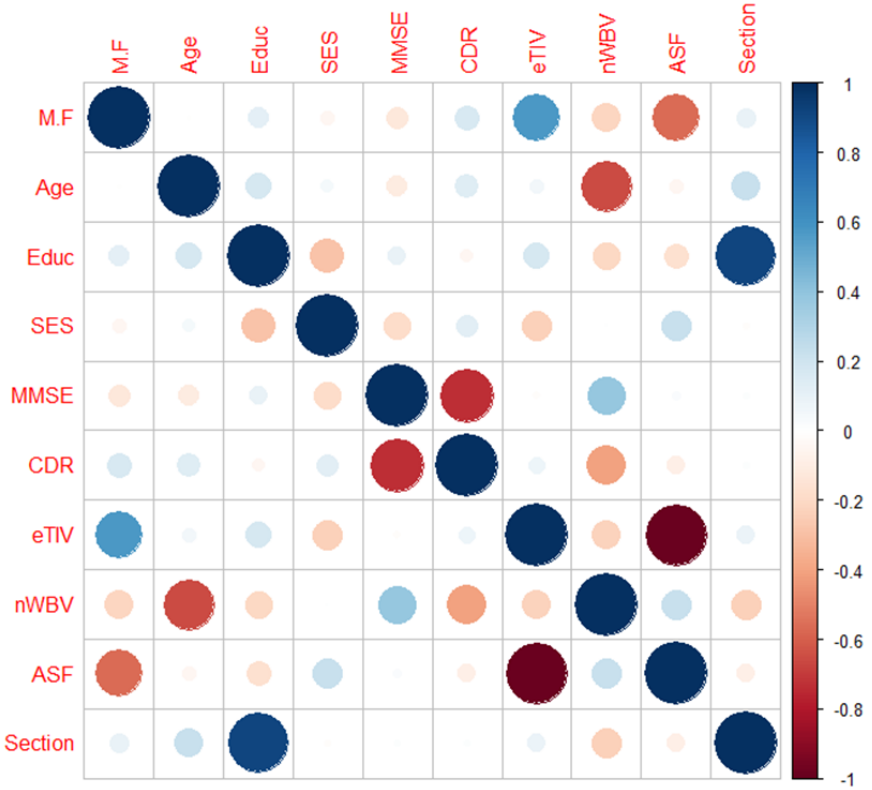

Figure 1: Correlogram

Based upon the results of the correlogram, I see that the variables that are correlated the most with *CDR* are *MMSE* and *nWBV*. It should be noted that both *MMSE* and *nWBV* are negatively correlated with *CDR*. The next three strongest are *M.F*, *Age*, and then *SES*. In order

to get a better understanding of the relationship between these variables and *CDR*, a series of scatterplots were created, pictured below:

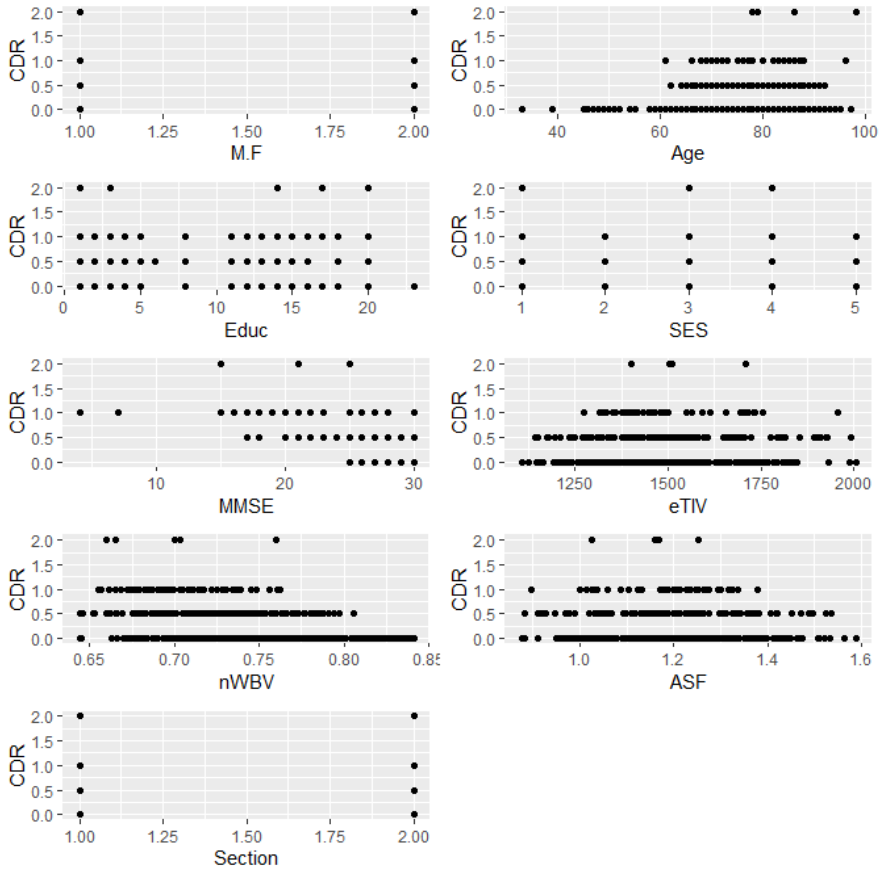

Figure 2: Scatter Plots

Overall looking at these scatterplots, it is immediately apparent that there is a lack of patients with a *CDR* above 1.0, which can be noted for future data gathering. In looking first at *MMSE,* the strongest negatively correlated variable to *CDR,* it can be seen there is quite a spread of values from 0 to 30, with a bulk of the patients scoring in the upper twenties. It logically makes sense that it is strongly negatively correlated with *CDR* since the higher patient scores on the *MMSE*, the better their mental state is, thus a higher *CDR* would correlate with a lower *MMSE*. However, it should be noted that since this is so strongly correlated, it would be ideal to have more patients scoring lower on the test in the future to determine if this relationship is truly as strong as our current data indicates. Next, we look at *nWBV*, the second strongest negatively correlated variable to *CDR*. Here we note that there is a good spread of *nWBV* among the levels of *CDR*, with the relationship again leaning moderately towards a negatively correlated one, as *nWBV* seems to increase, *CDR* seems to decrease. *SES* appears to have a good spread of data indicating good representation of different income levels of patients, suggesting this could be a sound variable for prediction. *Age* as well seems to cover a wide range of patients, however it would be nice to have more data points for the other levels of *CDR* in order to have a more robust idea about *Age*. Although it is hard to tell from the scatterplot, *M.F* does appear to be a bit lopsided with 349 patients being male and 221 patients being female, so it should be noted that there may be a bias for this variable.

(iii) Methods

Before testing with different models, the data was split into training and test sets. For the initial testing of the models, the data was split by saving every 5th observation (patient) as the test data and the rest of the observations as the training data. For the Monte Carlo Cross Validation later performed, the data was split by saving every 10th observation of the training and testing data as well, in order to get a better varying sense of the performance of each model.

In order to find a model that works well with the data given, the following methods were used to develop and test models with the data: LDA, Naive Bayes, Logistic Regression, KNN, LASSO, Ridge Regression, PCR, and PLS. LDA works by computing the means and covariance matrices of the data for each class, and then finding a linear combination of the features that maximizes the ratio of between-class variance to within-class variance. Naive Bayes works by first calculating the prior probabilities of each class based on the frequency of their occurrence in the training data. Then, it estimates the conditional probabilities of each feature given each class, again based on the training data. Finally, it uses Bayes' theorem to calculate the posterior probability of each class given a new observation, and selects the class with the highest probability as the predicted class. The logistic regression model estimates the parameters of the sigmoid function that best fit the training data, and then uses these parameters to make predictions on new data. The KNN algorithm works by finding the k closest training examples (neighbors) in the feature space to a new, unseen example and using the majority class or average value of the *k* neighbors as the prediction for the new example. For the initial runs of KNN, multiple models were run with *k* values from 1 - 50. Upon review of the error on the test set, $k = 1$ was selected for the cross validations. For each of these methods, they were all given the same Multinomial Model that was developed by finding the combination of variables that produced the lowest AIC score. This means that the final model was the one that explains the greatest amount of variation with the least possible number of variables. It was the following:

CDR = M.F + Age + Educ + MMSE + nWBV + ASF + Section

The rest of the methods were given a model that contained all of the variables to start.

Ridge Regression is a linear regression model with an added regularization term, which is the L2 regularization term. L2 helps to prevent overfitting, which discourages the model from assigning high weights to any one variable. This is similar to, but different from the next method LASSO. LASSO (Least Absolute Shrinkage and Selection Operator) is a linear regression method that adds a regularization term to the loss function, which is the L1 regularization term. L1 results in having less variables, since some of the variables have coefficients equal to zero, which removes them from the model. Principal Component Regression (PCR) is a regression technique that finds a set of eigenvectors of the variables, called principal components, that capture the most variation in the data. PCR was used to auto select the number of components through its use of cross validation and finding the minimum error. The regression coefficients for the principal components are estimated using ordinary least squares regression. For this dataset, cross validation showed that 9 principal components was the optimal choice. Like PCR, Partial Least Squares (PLS) uses principal components to represent the original variables, but it differs in that PLS specifically optimizes the relationships between the principal components and

the response variable, rather than just capturing the variation in the predictor variables. For this dataset, cross validation showed that 8 principal components was the optimal choice.

Once all the models were run and the error on the first iteration test set data was found, Monte Carlo Cross Validation with 100 repetitions was performed and the average errors on the second iteration test set was found to evaluate performance of the models.

(iv) Results

Upon creation of the eight models using the training data and testing them for their performance on the test data, they were further tested using Monte Carlo Cross-Validation 100 times. The following tables list the findings of these evaluations:

Model (Using Lowest AIC Score Formula)	**Multinomial Logistic Regression**	**KNN (k=11, chosen from 50 values tested)**	**LDA**	**Naïve Bayes**
Test Error	0.3070	0.2368	0.2807	0.3421
M.C. Test Error	0.2509	0.2733	0.2774	0.2523
M.C. Test Variance	0.0522	0.0558	0.0629	0.0569

Model (Starting with all variables)	**LASSO**	**Ridge Regression**	**PCR**	**PLS**
Test Error	0.0852	0.0872	0.0865	0.0865
M.C. Test Error	0.0653	0.0646	0.0659	0.0638
M.C. Test Variance	0.0005	0.0005	0.0006	0.0005

Since LASSO and PLS produced the lowest average test errors on average, I wanted to perform statistical difference testing against the other similar performing models, the results are as follows:

	LASSO	**Ridge Regression**	**PCR**	**PLS**
LASSO t.test p-value		0.001742	2.205e-08	0.002529
PLS t.test p-value	0.002529	6.299e-07	2.079e-10	

(v) Findings

Based upon the testing of all 8 models, the models developed using the LASSO and PLS methods were the best performing. Both of these models had the average lowest testing errors across the two iterations of test sets. However, in looking at the p-values of the t-tests comparing these two models to Ridge Regression and PCR, it is apparent that PLS performed statistically significantly better than both Ridge Regression and PCR. Meanwhile, LASSO only performed statistically significantly better than PCR. From this finding alone, I would be able to narrow my choice for continuing research with the model developed from the PLS method.

Interestingly, the testing errors from the methods that began with the Multinomial Model with the lowest AIC score all performed significantly worse than LASSO, Ridge Regression, PCR, and PLS. I suspect this for a couple of reasons. The first being that the data set that was used was certainly on the smaller side, and it was made even smaller when split into training and test sets. In the case of LDA, I know that the method sometimes does not work well with categorical variables, of which it was given two, *Section* and *M.F*. I suspect that these four models also had issues since I can see from the correlation plot that variables used in them would have had an issue with multicollinearity, the strongest of which being *Educ* and *Section*.

If I were to continue research and had the opportunity to collect more data, there would be several things to note. LASSO found that the most important variables were *M.F*, *Age*, *MMSE*, and *nWBV*. As more data is collected, I would find it very interesting to explore whether a patient is male or female truly has a significant impact on predicting *CDR*, or if it just happened to be that way for this particular dataset. As noted earlier, there is a lack of patients with a *CDR* above 1.0, and since the scale goes up to 3.0, it would be very beneficial to perhaps aim to have a better representation of patients with a *CDR* in that range of 1.5 to 3.0. Another consideration as the dataset grows in size would be to revisit the four modeling methods that did not work out so well for this current dataset. Perhaps with a larger set and picking variables in a different method other than AIC score will produce models that perform even better than our current top choices. To push the research even further, I would be interested in seeing more variables based upon the MRI scans themselves, much like the variable *nWBV*. Perhaps there is more prediction power from the scans than we currently realize. Overall, it is exciting that such great models were developed! Even more exciting is the potential of more research that can be done and more models that can be developed with the hope of bettering care around the globe!

(v.2) Lessons Learned

I can definitely say I learned a lot from this course. Certainly my R skills improved massively, and the introduction to the different modeling methods and the means in which to test those model predictions have been an incredibly useful set of skills I already use for the other course I am taking this semester in addition to a little at work. It has also been very helpful to learn the advantages and disadvantages of the different modeling methods, in addition to learning which situations and datasets are better with particular modeling methods. In terms of feedback on how to improve the course, I am a little bit conflicted. On one hand, I really enjoyed the open endedness along with the very responsive TAs on Piazza as it allowed me to really push myself to make a cohesive and informative project. But on the other hand, the open ended format of the instructions for both the written and presentation of the project gave me a little bit of anxiety by having me wonder if I did enough. Either way, it was not much of a problem and so

really the only feedback I have was that I really enjoyed this course and learned a good set of skills from it!

(vi) Bibliography

1. Creavin ST, Wisniewski S, Noel-Storr AH, et al. (January 2016). "Mini-Mental State Examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations" (PDF). The Cochrane Database of Systematic Reviews. 2016
2. "Differential diagnosis dementia". NICE. https://www.nice.org.uk/