Depression Patient SPECT Brain Scan Analysis

Client: Dr. Daniel Amen, Amen Clinics
Consultants: Maime Guan, Lingge Li, Duy Ngo, Fulya Ozcan, Dustin Pluta, Yuxiao Wang

January 22, 2017

Abstract

In the treatment of psychiatric and neurological disorders, determining an effective treatment for a specific individual patient is of paramount importance. In order to identify patients with a higher likelihood of response to a proposed treatment for clinical depression, data from 954 depression patients from Amen clinics were analyzed with logistic regression, random forest and other methods to predict treatment outcome. Random forests were the best performing of these methods achieved 56% accuracy in correctly classifying treatment responders from non-responders. Several brain regions are found to be statistically significant along with compliance indicator. The same methods are also able to separate depression patients from ADHD patients with AUC 0.72.

1 Introduction

Single photon emission computed tomography (SPECT) is a recent neuroimaging technique that has been applied to identify and understand various brain related health issues. SPECT shares some characteristics with other imaging techniques, particularly with Positron Emission Tomography (PET). Both SPECT and PET use radioactive tracers to produce reconstructed images of internal biology, which is capable of producing images with a high degree of detail. SPECT offers a reasonable alternative to PET, and can be cheaper and easier to use in practice due to employing a tracer with longer half-life and due to differences in the cost of scanners.

By producing a three dimensional map of blood flow activity in the brain, SPECT imaging can help identify abnormalities in a patient that may be related to a particular psychiatric affliction, for instance depression or Attention Deficit Hyperactivity Disroder. The Amen Clinics are one of the most prominent purveyors of the usefulness of SPECT imaging in the diagnosis and treatment of psychiatric patients, and have produced much work on a variety of diagnostic applications of SPECT. Previous results using SPECT include successful classificaition of NFL players from healthy subjects [1], identification of

significantly decreased perfusion in the prefrontal cortex among adolescent ADHD patients [2], and the classification of traumatic brain injury patients from post-traumatic stress order patients [3].

Additionally, SPECT has recently been shown to be clinically useful in the diganosis of Alzheimer's Disease using SPECT imaging data along with standard clinical measurements. A study by Pimlott and Ebmeier was [4] was able to correctly diagnose afflicted patients with over 80% accuracy.

Many of the patients at the Amen Clinics suffer from chronic depression or other psychological disorders. Among the wider population of psychiatric patients, the patients at the Amen Clinics often have a history finding little success with traditional or standard treatments for their disorders. For these cases, the physicians as the Amen Clinics will devise and prescribe therapies consisting of dietary and lifestyle changes as an alternative form of treatment. While potentially an effective and healthy treatment option, these dietary therapies may require many months before it can be determined whether there is a significant improvement in patient depression levels. Thus it is of interest to identify those patients with a higher likelihood of treatment response. Toward this end, the present analysis evaluates the performance of a variety of models for the prediction of treatment response for depression patients using clinical and SPECT data. In the absence of healthy controls, depression patients are also compared to ADHD patients without depression in an attempt to identify those brain regions which differ significantly in the depression patient population.

The remainder of the paper is organized as follows. Section 2 gives data description and exploratory analysis. Section 3 presents the results of a variety of machine learning models applied to the prediction of treatment response in depression patients. Following the broad results of these models, a more detailed presentation of the results of Bayesian logistic regression is provided in Section 4. Section 5 gives a comparative analysis of ADHD and depression patients using a random forest model. Finally, Section 6 summarizes the main points of the analysis and provides some discussion on possible directions for future work.

2 Data Description and Feature Reduction

The data set contains a total of 1103 patients with a Depression diagnosis and 1449 patients with ADHD. The main feature sets are: Demographics, Comorbidities, T_Baseline, T_Concentration, Max cluster size, and Min cluster size. There are two challenges with this data set. The first challenge is that many variables have missing values. Some variables

	Depression ($N = 1103$)	ADHD ($N = 1449$)
Responder	732 (0.66)	455 (0.31)
ΔBDI	-11.5 (12.29)	-4.29 (8.73)
Age	38.79 (17.26)	24.92 (16.73)
Gender		
Female	544 (0.563)	423 (0.437)
Male	559 (0.353)	1026 (0.647)

Table 1: Values are $Mean(SD)$ or $N(\%)$.

have more than 90% of their values missing. In order to be able to continue our analysis and reduce the noise, we eliminated the missing values in the following way. We first removed the columns that had more than 90% missing values, since they provide information for only a few patients in the data set. Then, we imputed the remaining missing values using Gelman's `mi` package in R, which imputes the missing values in an approximate Bayesian framework. This strong imputation technique allowed us to continue our analysis without losing any information. Some patients with an excessive number of missing values were removed, leaving 954 Depression patients for further analysis.

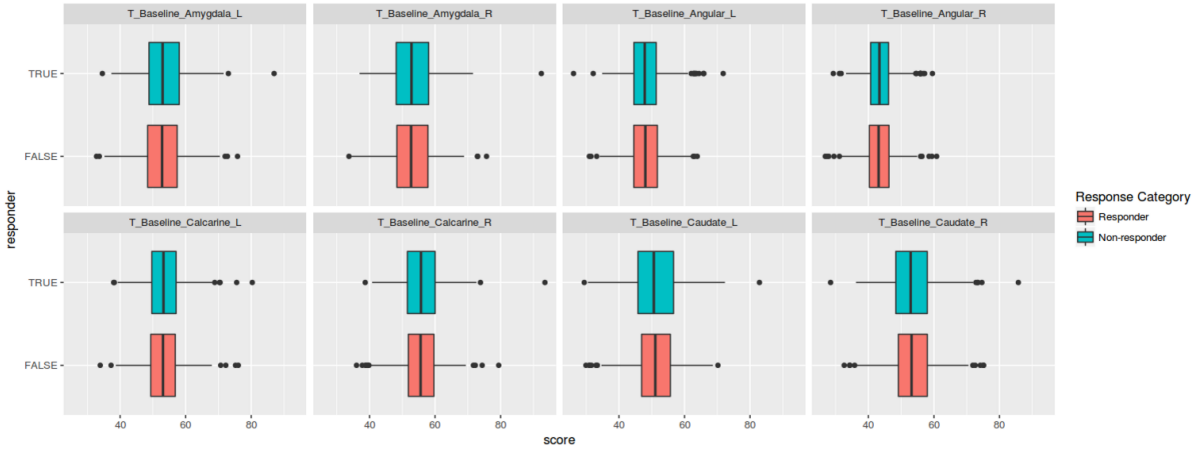

Figure 1: Boxplots of T scores in first eight brain regions colored by responder vs non-responder show minimal difference

The second challenge with this data set was that there were a large number of possible features that we could use. Included in the data were SPECT imaging measurements for 128 brain regions. SPECT data were collected under baseline (resting) condition and during a concentration task; from these measures t-normalized values and values scaled relative

to the global maximum measure. The present analysis is focused on the T-Baseline measures from the SPECT data. Recursive feature elimination was employed to select brain regions with T-Baseline feature are important in terms of their effect on co-diagnoses. This algorithm finds the important features by repeatedly constructing a model and removing the features that have low importance. We determined 10 important co-diagnoses which for depression, and found which brain regions were affecting them: Anxiety Disorder, Childhood Disorder, Attention Deficit Disruptive Behavior, ADHD, Substance Abuse Disorder, Temporal Dysfunction, Frontal Lobe Dysfunction, Diagnosed Brain Trauma, GAD, PTSD.

Anxiety Disorder
T_Baseline_Cerebellum_Crus2_R,T_Baseline_Parietal_Sup_R, T_Baseline_Temporal_Inf_Post_L, T_Baseline_Vermis_8, T_Baseline_Occipital_Inf_R, T_Baseline_Occipital_Inf_L, T_Baseline_Precentral_R,T_Baseline_Postcentral_R, T_Baseline_Cerebellum_Crus1_L, T_Baseline_Occipital_Sup_L
Childhood Disorder
T_Baseline_Frontal_Mid_Orb_R_10, T_Baseline_Cerebellum_7b_R, T_Baseline_Cerebellum_8_L, T_Baseline_Paracentral_Lobule_L, T_Baseline_Hippocampus_R, T_Baseline_Pallidum_L, T_Baseline_Putamen_L, T_Baseline_Frontal_Inf_Orb_R, T_Baseline_Temporal_Inf_Ant_R, T_Baseline_Amygdala_L
Attention Deficit Disruptive Behavior
T_Baseline_Cerebellum_7b_R, T_Baseline_Cerebellum_8_L, T_Baseline_Frontal_Mid_Orb_R_10, T_Baseline_Amygdala_L, T_Baseline_Paracentral_Lobule_L, T_Baseline_Frontal_Mid_Orb_L_9 T_Baseline_Pallidum_R, T_Baseline_Temporal_Inf_Post_R, T_Baseline_Caudate_L, T_Baseline_Temporal_Inf_Ant_R
ADHD
[T_Baseline_Frontal_Inf_Orb_R, T_Baseline_Frontal_Mid_Orb_R_10, T_Baseline_Cerebellum_7b_R, T_Baseline_Hippocampus_R, T_Baseline_Cerebellum_8_L, T_Baseline_Temporal_Mid_Ant_R, T_Baseline_Angular_R, T_Baseline_Hippocampus_L, T_Baseline_Paracentral_Lobule_L, T_Baseline_Cerebellum_Crus1_L
Substance Abuse Disorder
T_Baseline_Heschl_L, T_Baseline_Temporal_Sup_Post_R, T_Baseline_Paracentral_Lobule_L, T_Baseline_Hippocampus_L, T_Baseline_Calcarine_L, T_Baseline_SupraMarginal_R, T_Baseline_Temporal_Mid_Post_R, T_Baseline_Cerebellum_3_R, T_Baseline_Temporal_Inf_Ant_L, T_Baseline_Vermis_6
Temporal Dysfunction
T_Baseline_Amygdala_R, T_Baseline_Vermis_6, T_Baseline_Cingulum_Mid_R, T_Baseline_Lingual_L, T_Baseline_Insula_L, T_Baseline_Caudate_L
Frontal Lobe Dysfunction
T_Baseline_Cingulum_Ant_R, T_Baseline_Cingulum_Mid_R , T_Baseline_Cerebellum_Crus2_R, T_Baseline_Rectus_R, T_Baseline_Putamen_R, T_Baseline_Temporal_Inf_Ant_R, T_Baseline_Cingulum_Ant_L, T_Baseline_Cerebellum_Crus2_L, T_Baseline_Frontal_Mid_Orb_R
Diagnosed Brain Trauma
T_Baseline_Cerebellum_8_R, T_Baseline_Temporal_Mid_Ant_L, T_Baseline_Paracentral_Lobule_R, T_Baseline_Cingulum_Ant_L, T_Baseline_Amygdala_R , T_Baseline_Frontal_Sup_Medial_L, T_Baseline_Frontal_Sup_Orb_R, T_Baseline_Temporal_Sup_Ant_R, T_Baseline_Amygdala_L
GAD
T_Baseline_Frontal_Mid_R, T_Baseline_Temporal_Inf_Mid_L , T_Baseline_Occipital_Inf_L, T_Baseline_Cerebellum_Crus2_L, T_Baseline_Caudate_R, T_Baseline_Cerebellum_3_R, T_Baseline_Frontal_Sup_Medial_R, T_Baseline_Frontal_Sup_R, T_Baseline_Frontal_Mid_Orb_R_10, T_Baseline_Occipital_Sup_L
PTSD
T_Baseline_Cingulum_Ant_R, T_Baseline_Frontal_Inf_Oper_L, T_Baseline_Supp_Motor_Area_L, T_Baseline_Cerebellum_Crus1_R, T_Baseline_Temporal_Inf_Mid_R, T_Baseline_Temporal_Pole_Sup_R, T_Baseline_Temporal_Mid_Ant_R , T_Baseline_Parietal_Inf_R , T_Baseline_Supp_Motor_Area_R, T_Baseline_Temporal_Inf_Ant_L

Table 2: Important features for co-diagnoses

As a result of the recursive feature elimination, we identified 10 regions for each co-diagnoses. After identifying important regions for co-diagnoses, we also identified significant brain regions that affect the outcome using logistic regression. The 9 significant regions at 5% level from logistic regression were: Cerebellum_9_L, Cingulum_Post_L, Fusiform_L, Fusiform_R, Pallidum_L, SupraMarginal_L, Temporal_Mid_Post_L, Temporal_Mid_Post_R, Temporal_Pole_Sup_L,Vermis_3.

In this section, we consider the problem of predicting the response (yes or no) based on the characteristics of the patients. In this work, we mainly focus on the t_baseline features and the t_concentration features, both of which are measures of the brain image in different regions.

3 Classification models

There are a wide variety of machine learning models in current use, each with unique advantages and weaknesses. The performance of a model will depend largely on the specific characteristics of the data, such as the dimensionality and noise, and on the underlying structure of the observations, such as whether there are naturally forming clusters in the observations. Thus there is no single model that can work well for all datasets. In this work, a number of different classification algorithms were evaluated in order to find the ones that are most suitable for this problem. Specifically, the following classification models were considered.

- k Nearest Neighbors (kNN): the prediction is based on the majority vote by labels of the neighboring training data. Works well when the dimensionality of the data is not too large, and when the data naturally separate into a few distinct clusters in the original sample space. Performance can suffer greatly from high-dimensionality and when boundaries between clusters are irregular.
- Support Vector Machine (SVM): a discriminative classifier defined by a separating hyperplane, where the margin of the training data is maximized. Multiple kernels can be used to incorporate features of more complex structures. Popular kernels include linear kernel, polynomial kernel and RBF(Gaussian) kernel. SVMs are flexible and powerful classifiers, but it difficult to interpret the results in terms of the original features.
- Gaussian Process: a classifier based on Gaussian process. Can be computationally intensive, and may require a substantial amount of data to achieve a good fit.
- Decision Tree: an algorithm that recursively breaks down the training set into smaller pieces according to one feature at each time, and thus constructing a tree-structured model, where the prediction is computed via the leaf nodes. Tend to be sensitive to overfitting and tuning parameters.

- Random Forest: ensemble of decision trees that are constructed using a subset of the features and training data. It usually has better performance comparing to decision tree as it reduces the possibility of model overfitting. Interpretation of important features and resulting decision boundaries can be difficult. Requires some model tuning
- Multilayer Perceptron: a feedforward artificial neural network model that maps the input data to output through multiple hidden layers. Can produce very good predictive accuracy, but the required amount of data may be prohibitive, and there is no standard method to identify and interpret important features.
- AdaBoost: uses a sequence of weak learners that are tweaked in favor of the error made by the previous learners.
- Naive Bayes: a simple Bayesian classifier where the features are assumed to be independent conditional on the classification labels.
- Quadratic Discriminant Analysis (QDA): the algorithm is similar to linear discriminant analysis (LDA) but there are no assumptions on the covariance matrices.

Because of the complex and high-dimensional nature of the data, and the expectation of complicated decision boundaries, the it was anticipated that the random forest and SVM models would show the best performance, while kNN, Naive Bayes, and QDA were not expected to perform well. A comprehensive treatment of the above classification models are given in [5] and [6].

To construct the classifiers, the dataset was randomly split into training and testing data, where the testing data contains about 20% of the data while the training data consists of the rest 80%. The model training was purely based on the training data and the evaluation was performed on the testing data. The performance of the model was evaluated by the overall classificaiton accuracy, which is the proportion of the testing data for which the model gives the correct prediction.

Table 2 displays the classification accuracy for the various algorithms. All of the algorithms gave accuracy around 50%, which means all the models are no better than random guessing. Considering that the labels are not balanced, the models actually perform worse than just predicting "Responder" for all test samples. Figure 2 shows the decision boundary for the models, plotted on a 2-D space spanned by the first two principal components. The projection onto the space spanned by the first two principal components accounts for 70% of the total variation. It is observable that the two classes are overlapping and are not linearly separable.

Though disappointing, it does not necessarily mean that the models are useless for this problem, as the results are obtained for single model that are trained on raw features (no

transformations at all). Better feature engineering, model stacking and ensembling could boost the performance, when there are enough training samples. Moreover, the output of the models can be informative, for example, random forest model can also rank the importance of the input features. It should also be noted that the training data here consists solely of patients with depression. Considering the noise and complexity of the data, a more powerful and precise model could possibly be achieve through the inclusion of a control group of non-depression patients.

Model Name	only T_Baseline	Only T_Concentration	Both
Nearest Neighbors	49.2%	51.8%	49.2%
Linear SVM	48.2%	52.9%	48.2%
RBF SVM	52.4%	52.4%	52.4%
Gaussian Process	48.2%	51.8%	51.8%
Decision Tree	55.5%	46.1%	51.3%
Random Forest	51.3%	54.5%	53.9%
Multilayer Perceptron	48.7%	50.3%	48.2%
AdaBoost	46.6%	47.6%	51.3%
Naive Bayes	49.2%	46.6%	45.5%
Quadratic Discriminant Analysis	53.9%	56.0%	50.8%

Table 3: Comparison of the classification accuracy of different models. Decision trees, random forests, and QDA perform the best overall, although none of the models does sufficiently well for clinical application, rather, the results here do not appear to be significantly better than random guessing overall.

4 Bayesian Logistic Regression

To investigate the relationship between the SPECT image scans and some other covariates of interest (e.g., anxiety disorder, diagnosed brain trauma, PTSD) with the outcome of response to treatment, we fit three different logistic regression models to the data in a Bayesian framework using 'MCMCpack', software that facilitates MCMC-based computational Bayesian inference [7].

Logistic regression models have three components: the outcome variable, the logistic function link, and the predictors we choose to include in the model. In all three models, the outcome variable is whether or not patients responded to treatment. We consider different predictors to be included in the three different models, explained later in this section. In the Bayesian logistic regression framework, there is an additional component of incorporating prior knowledge into our model. This is done by setting priors for the coefficients of the predictors in the models. We chose to use uninformative priors for the coefficients of

predictors in all three models in our analyses because we have no prior information from existing literature on how these covariates might relate to response to hollistic treatment of depression. However, to control for pre-BDI scores, we naturally include the pre-BDI scores of each patient as a predictor in all three of our models.

For our first logistic regression model, we only consider pre-BDI scores and the top eight baseline brain regions from the dimensionality-reduction analyses: left cerebellum 7B, right crus cerebellum 1, left superior temporal pole, right rectus, right cerebellum 9, right superior temporal lobe, vermis 3, and right superior occipital. The model equation 1 is given below, along with the estimates of each coefficient using exponentiated posterior sample means and 95% credible intervals.

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_9 X_{9i} \tag{1}$$

where $X_{1i} \ldots X_{9i}$ are pre-BDI and the top 8 brain regions.

	exp(means)	exp(95% credible intervals)
Intercept	0.312	(0.049, 2.021)
Pre-BDI	1.009	(0.990, 1.030)
Left cerebellum 7B	0.980	(0.944, 1.018)
Right crus cerebellum 1	1.030	(0.989, 1.027)
Left superior temporal pole	1.015	(0.949, 1.088)
Right rectus	1.020	(0.980, 1.061)
Right cerebellum 9	0.990	(0.957, 1.024)
Right superior temporal pole	1.029	(0.969, 1.094)
Vermis 3	**0.964**	**(0.936, 0.993)**
Right superior occipital	1.002	(0.960, 1.047)

Table 4: Summary of results from Model 1.

The results of our Bayesian logistic regression analysis indicate that the only brain region out of the eight regions tested in our model that is weakly associated with odds of responding to treatment is Vermis 3. We estimate that each additional unit of baseline activation of Vermis 3 is associated with about a 3.6% *decrease* in the odds of response to treatment, after controlling for pre-BDI scores and the baseline activation of the other 7 brain regions. Furthermore, there is a 95% probability that this association of Vermis 3 with odds of response to treatment is between a 0.7% and 6.4% decrease.

In Model 2, we consider pre-BDI scores with some co-diagnoses of interest: anxiety disorder, ADHD, substance abuse disorder, frontal lobe dysfunction, diagnosed brain trauma,

and PTSD. The equation for Model 2 is given below, along with results from our Bayesian logistic regression analysis.

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_7 X_{7i} \tag{2}$$

where $X_{1i} \ldots X_{7i}$ are pre-BDI and the 6 co-diagnoses of interest.

	exp(means)	exp(95% credible intervals)
Intercept	1.297	(0.580, 2.471)
Pre-BDI	1.010	(0.992, 1.032)
Anxiety Disorder	**0.613**	**(0.417, 0.957)**
ADHD	0.863	(0.625, 1.195)
Substance Abuse Disorder	1.337	(0.854, 2.195)
Frontal Lobe Dysfunction	0.872	(0.635, 1.250)
Diagnosed Brain Trauma	0.879	(0.631, 1.247)
PTSD	0.964	(0.642, 1.427)

Table 5: Summary of results from Model 2.

From these results, the only co-diagnosis out of the six that are tested in Model 2 that is associated with odds of response to treatment is anxiety disorder. After adjusting for the other 5 co-diagnoses, we estimate that being diagnosed with anxiety disorder in addition to depression is associated with about a 39% *decrease* in the odds of response to treatment, compared to not being diagnosed with anxiety. Furthermore, there is a 95% probability that being diagnosed with anxiety is associated with between a 4.3% and 58.3% decrease in odds of response to treatment.

Lastly, we consider a third and final Bayesian logistic regression model in which we include pre-BDI scores, the eight brain regions from Model 1, the six co-diagnoses from Model 2, compliance ratings, age, and gender. The equation of Model 3 is given below, along with the table of results from our Bayesian logistic regression analysis.

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_{20} X_{20i} \tag{3}$$

where $X_{1i} \ldots X_{20i}$ are pre-BDI, the top 8 brain regions, 6 co-diagnoses of interest, compliance ratings, age, and gender.

Consistent with findings from Models 1 and 2, the results of our Bayesian logistic regression analysis of Model 3 show that Vermis 3 and anxiety disorder diagnosis are associated with the odds of response to treatment. After controlling all of the other covariates in

	exp(means)	exp(95% credible intervals)
Intercept	0.343	(0.069, 3.057)
Pre-BDI	1.014	(0.992, 1.033)
Left cerebellum 7B	0.980	(0.941, 1.023)
Right crus cerebellum 1	1.038	(0.998, 1.077)
Left superior temporal pole	1.024	(0.956, 1.101)
Right cerebellum 9	0.991	(0.957, 1.024)
Right superior temporal pole	1.027	(0.966, 1.089)
Vermis 3	**0.959**	**(0.935, 0.984)**
Right superior occipital	1.000	(0.960, 1.048)
Anxiety Disorder	**0.638**	**(0.443, 0.922)**
ADHD	0.921	(0.705, 1.277)
Substance Abuse Disorder	1.403	(0.974, 1.980)
Frontal Lobe Dysfunction	0.819	(0.581, 1.169)
Diagnosed Brain Trauma	0.922	(0.665, 1.260)
PTSD	1.052	(0.777, 1.500)
Compliance: Not	0.533	(0.271, 1.109)
Compliance: Somewhat	0.837	(0.499, 1.349)
Compliance: Very	**2.131**	**(1.262, 3.585)**
Age	1.002	(0.990, 1.013)
Gender: Male	1.186	(0.830, 1.692)

Table 6: Summary of results from Model 3.

the model, we estimate that each additional unit of the baseline activation of Vermis 3 is associated with about a 4.1% *decrease* in the odds of response to treatment. We also find a negative association between anxiety disorder and odds of response to treatment. Adjusting for all other covariates in the model, we estimate that being diagnosed with anxiety disorder is associated with about a 36.2% decrease in odds of response to treatment. Or in other words, patients who also have anxiety in addition to depression are less likely to respond to hollistic treatment of depression, when compared with patients who are only diagnosed with depression.

A new finding from the analysis of Model 3 indicates that holding all other covariates constant, a self-report of being "very compliant" with the treatment plan is associated with about a 113% increase in the odds of response to treatment, compared to not giving a self-report on compliance to treatment. Furthermore, we find that age and gender are not associated with odds of response to treatment.

All of the model results are based on 4 chains of 10,000 samples samples each. The chains were verified for convergence using the standard $\hat{R}$ statistic [8].

5 Bayes Factors

In addition to estimating the coefficients of our models in a Bayesian framework, we can also quantify the evidence for each of our models by making pairwise comparisons between them using Bayes factors. Bayes factors are a standard Bayesian approach to model selection that implicitly controls for goodness-of-fit and model complexity. It is the degree of change from prior to posterior information given by the data. We compute the natural logarithm of the Bayes factors between each pairwise comparison of our three models. A positive log Bayes factor indicates evidence for the first model in the comparison, and a negative log Bayes factor indicates evidence for the second model. Standard interpretive boundaries at log-odds of 2, 6, and 10 correspond to "moderate", "strong", and "very strong" (Kass & Raftery, 1995). The table below shows the matrix of the natural log Bayes factors resulting from all 9 pairwise comparisons between Models 1, 2, and 3.

	Model 1	Model 2	Model 3
Model 1	0.0	-29.4	55.3
Model 2	29.4	0.0	84.8
Model 3	-55.3	-84.8	0.0

Table 7: Log Bayes factors between all 3 models. Positive log Bayes factors indicate evidence for that model, while negative log Bayes factors indicate evidence against that model. Model 1: pre-BDI and top 8 brain regions from dimensionality-reduction analyses. Model 2: pre-BDI and 6 co-diagnoses of interest. Model 3: pre-BDI, the top 8 brain regions, and the 6 co-diagnoses of interest. As evident here, Model 3 performs the best.

There is overwhelming evidence that Model 2 with just pre-BDI and 6 co-diagnoses is preferred over both Models 1 and 3. In particular, the saturated Model 3 with the most predictors is the worst because it overfits the response data. Model 1 with just pre-BDI and 8 brain regions is preferred over the saturated Model 3, but not preferred when compared to Model 2. These results indicate very strong evidence that baseline activation of the 8 brain regions do not provide more information if pre-BDI and co-diagnoses are already included in the model. Or, in other words, knowing the pre-BDI scores and the 6 co-diagnoses is adequate; the SPECT readings of brain regions do not help much to further improve the model.

6 ADHD/depression Classification

Along with the Depression patient data set, data from 1449 ADHD patients were also available from the Amen Clinics records. These patients also had their SPECT data recorded, along with their Pre- and Post-BDI scores (even though they were not being treated for depression specifically). Since the primary focus of this study is on the change in BDI for Depression patients, all patients with a Depression diagnosis were included in the Depression patient set; some of the Depression patients have an ADHD diagnosis, but none of the patients in the ADHD set were diagnosed with depression.

In an effort to understand the association of SPECT readings with depression, a random forest model is employed to classify patients by depression diagnosis, using all SPECT baseline and concentration readings, and adjustment variables age and gender. The random forest model is constructed as an aggregation of decision trees, each fit to a randomly chosen subset of the available predictors, and using the Gini impurity to choose the best splits. For each individual tree constructed for the forest, 50 covariates are randomly selected from the 258 total covariates (128 T Baseline, 128 T Concentration, Age, and Gender); 1000 such trees are built, and their individual predictions are aggregated to produce the random forest's prediction.

First considering the random forest model for diagnostic classification of ADHD from Depression patients using only SPECT, Age, and Gender, the test set ROC, which has a AUC of 0.7153, indicating moderate overall predictive performance, but probably under the level of clinical applicability. The plot of variable importance is given below. The plot makes it clear that age is by far the strongest predictor of depression diagnosis, which is expected from the strong correlation between age and depression status in the sample. The second strongest predictor is the SPECT baseline reading for Vermis 1-2, with a larger importance score than even gender, which is known to be significantly associated with depression diagnosis. The variable importance plot suggests that a small number of the SPECT regions are strongly associated with depression diagnosis, with the majority of regions not providing significant information regarding depression status. A sparse signal is expected, and this result confirms the need for dimensionality reduction when modeling with SPECT data.

7 Conclusion

Our models did not predict patient outcome at an accuracy level that we had hoped for. There are several possible explanations. First of all, patients were categorized into two groups in a somewhat arbitrary manner. It is easy to imagine that there are not substantial differences between patients whose BDI changes slightly differ. Yet patients with

similar BDI changes could be responders and non-responders. Secondly, we did not expect logistic regression to achieve satisfactory results as the data are not linearly separable. In other words, responders and non-responders have comparable brain region scores overall. The only brain region of those considered to have some statistically significant association with the odds of respnose to treatment was the Cerebellar Vermis, for which increased blood flow was associated with a decrease in the odds of treatment response.

The machine learning models implemented here did show some success in separating depression and ADHD patients, and found brain regions adjusting for other variables. Notably, the most important brain regions were again associated with teh Vermis. It implies that SPECT scans provide useful insights beyond the demographic and diagnostic information. This, without a doubt, could support the effectiveness of using SPECT scan for patient diagnosis as several previous studies. It is worth noting that in the combined depression and ADHD dataset, many patients are diagnosed with both, which increases the difficulty of the classification task.

Previous studies have found evidence that the Vermis blood flow abnormalities may be associated clinical depression and other related cognitive disorders [9], [10], [11], thus this region may be an important target for future work. However, it should be noted that these results should be cautiously interpreted, as the large number of variables in the present data, many of which are noisy and contain some redundancy, pose a modelling challenge. In order to identify the most important features, many different models and strategies were attempted, thus these results should be understood from an exploratory perspective. While the present results may be suggestive, further work, ideally in the form of controlled, randomized experiments, would be needed to validate these findings and more precisely characterize the role of the Vermis in depression and the response to treatment.

In the search for biomarkers to identify the likelihood of positive treatment response, the above analysis suggests that, due to the inherent complexities of the disorders considered, a multi-modal approach to diagnosis is likely necessary. While SPECT readings may help predict treatment response, a wide range of clinical data for patients is probably necessary to build a clinically practical classification model. It should also be noted that the patient population of the Amen Clinics presents some unique challenges, in particular, the large number of co-diagnoses among the patients, and the general severity of the cases relative to the broader population of psychiatric patients. Because of this, additional SPECT readings on healthy subjects or depression patients from a broader population may be helpful in the further identification of useful biomarkers.

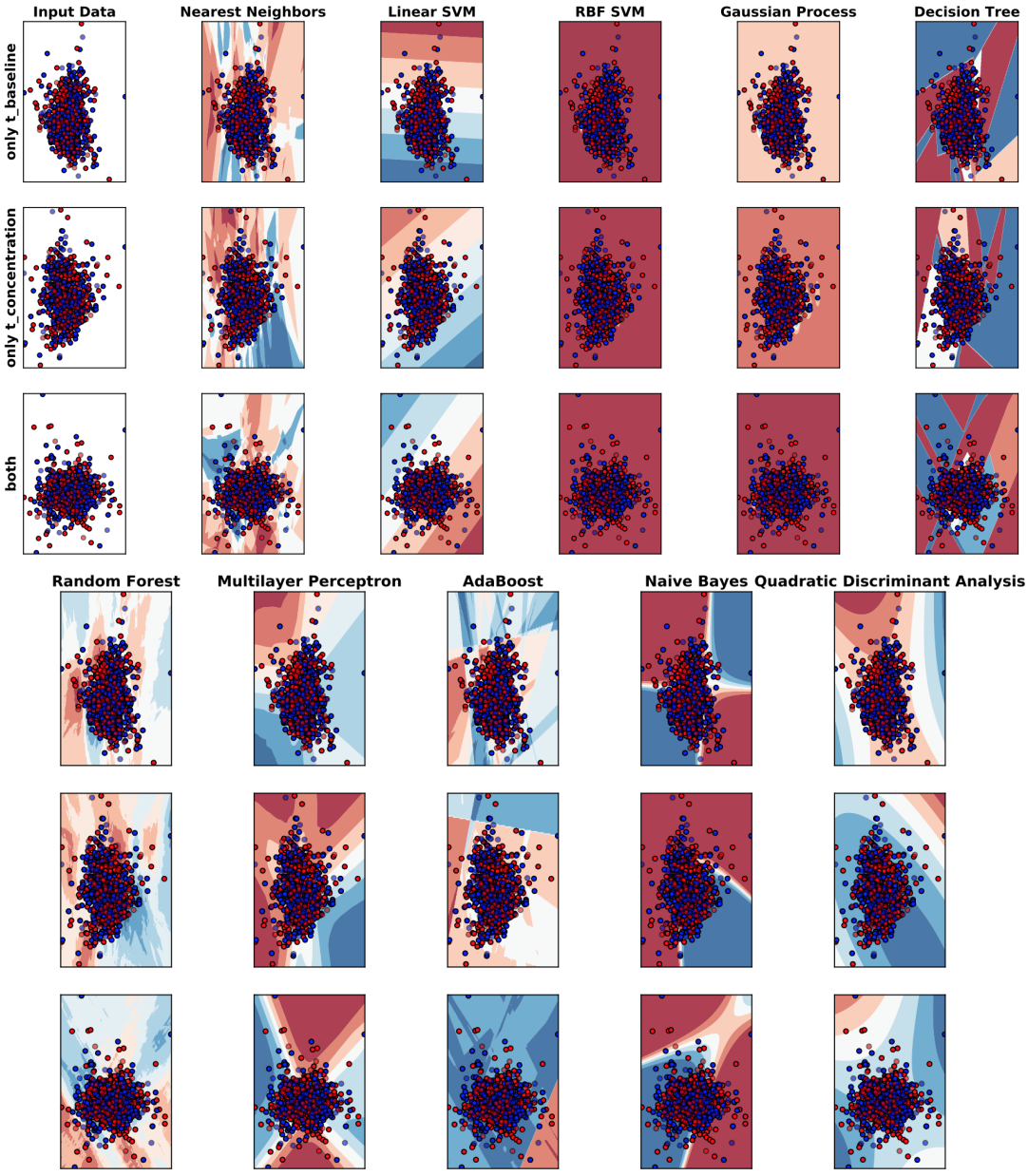

Figure 2: Decision boundary for the classifiers. The plots show a projection of the data into a lower dimensional space to allow for the visualization of the calculated decision boundaries. In an ideal setting, the red and blue points (responder and non-responder) would be separated along the decision boundaries. However, none of the models gives a particularly clear separation of the classes, suggesting that the predictors are not sufficiently related to the outcome, or that further denoising procedures need to be applied and new features extracted from the data.

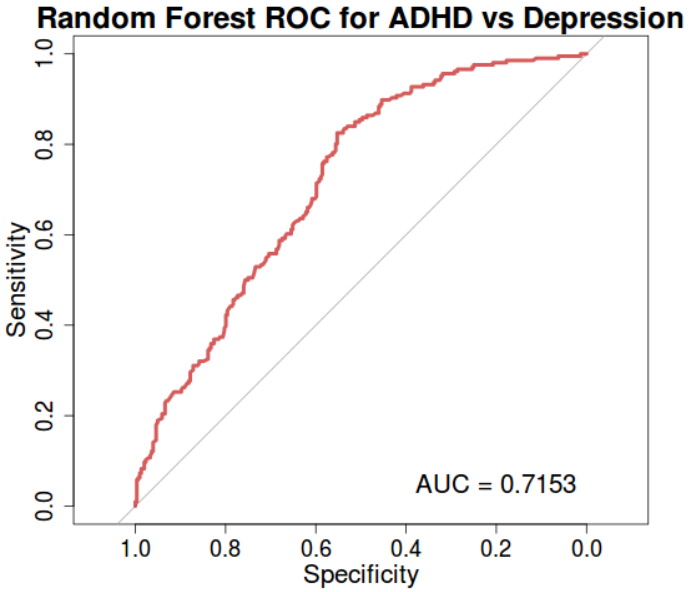

Figure 3: Test set ROC for random forest classification model for depression diagnosis. The predictors in this model are all SPECT baseline and concentration readings, and age and gender adjustment variables. The AUC of 0.7153 indicates that this model performs reasonably well overall in terms of varying levels of specificity and sensitivity. While the accuracy obtainable with the present model is likely not sufficient for clinical application, these results suggest that re-training the model with additional data may yield a practical model.

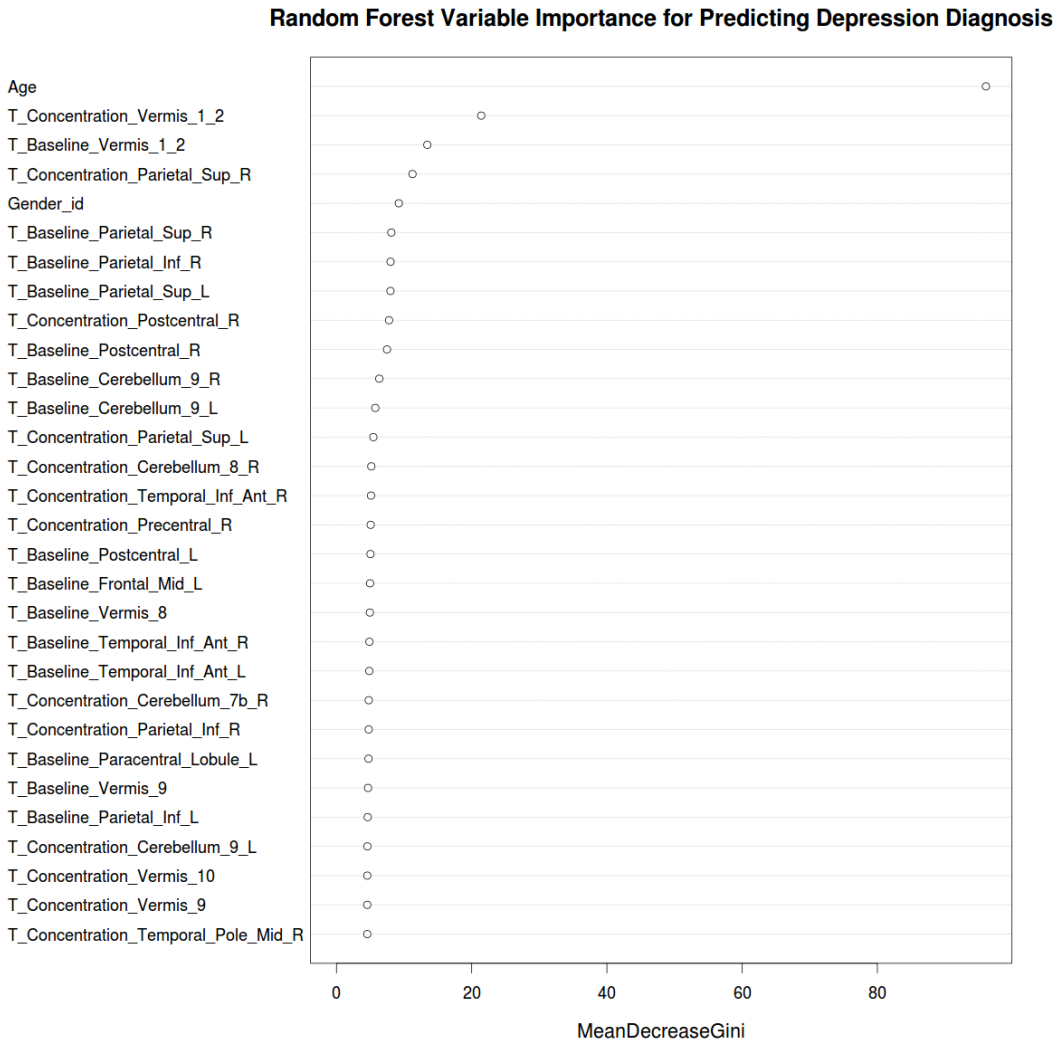

Figure 4: Variable importance plot for classifying depression patients from ADHD. Age the most important predictor by a wide margin. The most important SPECT co- variate is the baseline for Vermis 1-2.

References

[1] Daniel G Amen, Kristen Willeumier, Bennet Omalu, Andrew Newberg, Cauligi Raghavendra, and Cyrus A Raji. Perfusion neuroimaging abnormalities alone distinguish national football league players from a healthy population. *Journal of Alzheimer's Disease*, 53(1):237–241, 2016.

[2] Daniel G Amen and Blake D Carmichael. High-resolution brain spect imaging in adhd. *Annals of Clinical Psychiatry*, 9(2):81–86, 1997.

[3] Cyrus A Raji, Kristen Willeumier, Derek Taylor, Robert Tarzwell, Andrew Newberg, Theodore A Henderson, and Daniel G Amen. Functional neuroimaging with default mode network regions distinguishes ptsd from tbi in a military veteran population. *Brain imaging and behavior*, 9(3):527–534, 2015.

[4] SL Pimlott and KP Ebmeier. Spect imaging in dementia. *The British journal of radiology*, 2014.

[5] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics Springer, Berlin, 2001.

[6] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[7] Andrew D Martin, Kevin M Quinn, Jong Hee Park, and Maintainer Jong Hee Park. Package mcmcpack, 2016.

[8] Stephen Brooks. Markov chain monte carlo method and its application. *Journal of the royal statistical society: series D (the Statistician)*, 47(1):69–100, 1998.

[9] Christopher J Bench, Karl J Friston, Richard G Brown, Lynette C Scott, Richard SJ Frackowiak, and Raymond J Dolan. The anatomy of melancholia–focal abnormalities of cerebral blood flow in major depression. *Psychological medicine*, 22(03):607–615, 1992.

[10] SA Shah, PM Doraiswamy, MM Husain, PR Escalona, C Na, GS Figiel, LJ Patterson, EH Ellinwood, WM McDonald, OB Boyko, et al. Posterior fossa abnormalities in major depression: a controlled magnetic resonance imaging study. *Acta Psychiatrica Scandinavica*, 85(6):474–479, 1992.

[11] Dennis JLG Schutter and Jack Van Honk. The cerebellum on the rise in human emotion. *The Cerebellum*, 4(4):290–294, 2005.

[12] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[13] Joseph R. Phillis, Doaa Hewedi, Abeer Eissa, and Ahmed Moustafa. The cerebellum and psychiatric disorders. *Frontiers in Public Health*, 3(66):1–8, may 2015.

[14] K Yucel, A Nazarov, VH Taylor, K Macdonald, GB Hall, and GM Macqueen. Cerebellar vermis volume in major depressive disorder. *Brain Structure and Function*, 218(4):851–8, jul 2013.

[15] Karen Lansing, Daniel G Amen, Chris Hanks, and Lisa Rudy. High–resolution brain spect imaging and eye movement desensitization and reprocessing in police officers with ptsd. *The Journal of neuropsychiatry and clinical neurosciences*, 17(4):526–532, 2005.

[16] Daniel G Amen, Joseph C Wu, Derek Taylor, and Kristen Willeumier. Reversing brain damage in former nfl players: implications for traumatic brain injury and substance abuse rehabilitation. *Journal of psychoactive drugs*, 43(1):1–5, 2011.