Structural and Functional Brain Networks: From Connectome to Major Depression

Ling Cai, Zhaoqi Liu, Ximing Wen, Wan Zhang

May 10, 2021

1 Introduction

Connectome refers to the concept of the human brain as a large-scale complex network.[1] In other words, the human connectome captures the functional and structural connections of brain regions in terms of brain networks. The Functional Connectome (**FC**) refers to the network of correlated brain regions and the Structural Connectome (**SC**) refers to the connections of anatomical white matter in the brain.[1] With the development of network science and brain imaging techniques, we are able to extract the functional and structural connectome and study how the connections may vary for individuals in different groups.

Major depression is a psychiatric disorder, which is characterized by persistent low-mood and often accompanied by several heterogeneous symptoms for different individuals.[7] As a result, so far, the diagnosis of major depression is based on the person's self-assessment and a mental status examination, but no biological test can confirm it. Even though the pathophysiology of major depression is little known, the current studies indicate that the pathology may be distributed across many brain regions and circuits. For example, the structural and functional abnormalities in the amygdala, hippocampus and thalamus, which belongs to the subcortical limbic brain regions, have been found in depression patients.[6]

Major depressive disease has high prevalence, early onset and a persistent nature which affects millions of people worldwide. Identifying neuroimaging-based biomarkers will facilitate early risk detection, stratification and helping monitoring of interventions of this disease. In this report, we are interested in gaining the sharper insight into the relationship between brain connectivity and major depression. Firstly, we want to check whether the major depressive disorder will make changes in the brain networks. We use both the connectivity matrices and dimension-reduction PCA scores to perform tests and do visualization. Secondly, we try to fit models to predict the major depression using the brain connectome data. Our code and data can be downloaded from `https://github.com/Zhaoqi-0903/STOR893-Final-Project`

2 Method

2.1 Data Set

The data was obtained from the Human Connectome Project (HCP). 1206 subjects participated in the project. We used the extracted Functional connectome (FC) and Structural connectome (SC) data from Professor Zhang's research.[8] According to the Desikan-Killiany atlas, the human brain is parcellated into 68 cortical surface regions.[3] Thus, we have two 68 x 68 connectivity matrices for each individual, which measures the connected fiber counts across the 68 regions for both FC and SC. Since the high dimensional data are difficult for modeling, we used the tensor PCA scores from Professor Zhang, who used the tensor PCA (TNPCA) approach approximated the brain tensor network using 60 components, with the components ordered to have decreasing impact similarly to common practice in PCA. Thus, 1065 subjects each have 60 tensor PCA scores for FC and 60 tensor PCA scores for SC. The details of the method and result can be found here `https://github.com/zhengwu/TensorNetwork_PCA`

Besides the brain connectivity data, we also have 175 traits, which include the participant's cognition, motion, psychiatric history and so on. The two traits we explored here are:

- Trait 130: Whether the subject has experienced a diagnosed major depressive episode over his/her lifetime.
- Trait 131: The number of depressive symptoms endorsed for major depression the subject has.

The first trait of interest is a binary variable. The second trait of interest is a continuous variable which ranges from 0 to 9. In this report, we use connectivity matrices for circular plots and use tensor PCA scores for modeling.

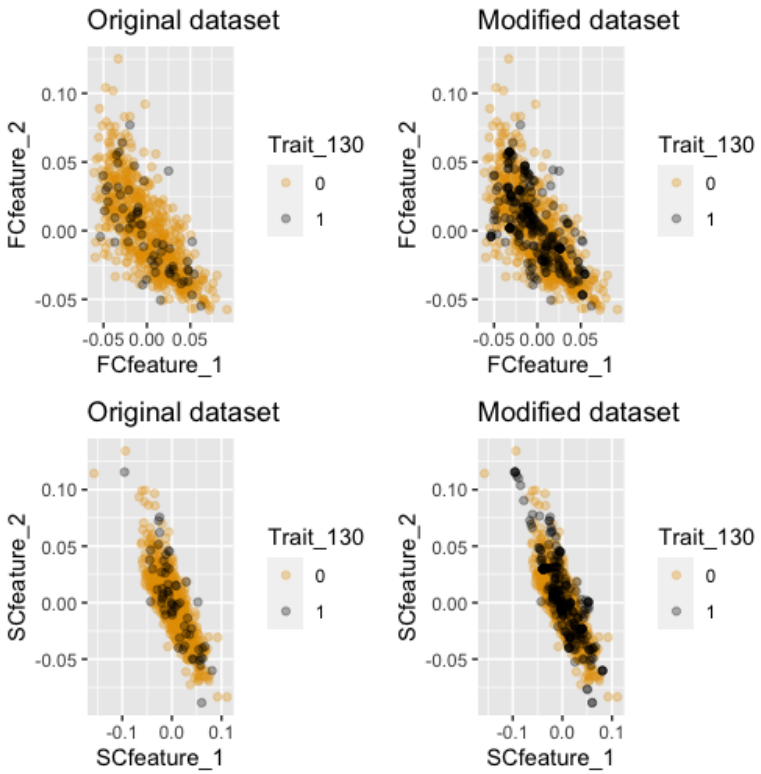


Figure 1: Comparison of original training set and oversampling training set. The upper shows the first two PC scores for FC and the bottom shows the first two PC scores for SC. The synthetic data seem to follow the same distribution as the original data for both FC and SC.

During modeling, we face the problem of imbalanced data set. For the total

1028 subjects after data cleaning, we only have 95 subjects have experienced major depressive symptoms over their lifetime. Such imbalanced data can make the classifier less sensitive and reduce the prediction accuracy. So we deal with the imbalanced data problem by using oversampling method. We first split the data into training and test set by a ratio of 2:1, and then generated 200 synthetic samples in the training set using synthetic minority oversampling technique (SMOTE). The data comparison can be seen in Figure 1. The plots indicate that the synthetic data follow the same distribution as the original data

2.2 Modelling

This section illustrates the models we use for prediction analysis. To predict Trait 130, which is a binary variable, we perform classification models including Logistic Regression, LDA, Naive Bayes, Elastic Net, SVM, Decision Tree and Random Forest. Trait 131 is the number of depression symptoms, which can be viewed as a continuous variable. So we adopt ridge regression, lasso, SVM, Regression Tree and Random forest for regression analysis.

2.2.1 Logistic Regression, LDA and Naïve Bayes

Logistic Regression Logistic Regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable. The logistic regression model itself simply models probability of output in terms of input and does not perform statistical classification (it is not a classifier), though it can be used to make a classifier, for instance, by choosing a cutoff value and classifying inputs with probability greater than the cutoff as one class, below the cutoff as the other.

LDA Linear Discriminant Analysis (LDA) is a dimensional reduction and supervised classification technique. It assumes that features in each class follow multivariate normal with different means but shared covariance. It works well when sample size is small and the Gaussian assumption holds.

Naïve Bayes Naïve Bayes classifiers are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong independence assumptions between the features. They are among the simplest Bayesian network models, but coupled with kernel density estimation and can achieve higher accuracy levels.

2.2.2 Ridge Regression, Lasso and Elastic Net

Since the TN-PCA data includes 120 FC and SC features, we adopt penalized regression to solve the high dimensional problem. The first two models come into our mind is the ridge regression and lasso with the penalized term ℓ_2 and ℓ_1. The ridge regression keeps all the predictors, while lasso naturally eliminates some predictors. The problem with lasso is that, sometimes it shrinks all the coefficients to 0. The probable reason is that we have an unbalanced data with less than 10% positive samples, and it gets much better after we prepare the data with oversampling. But still we also try the generalized Elastic Net, which is a mixture of ridge regression and lasso, so that we never get an empty model.

The elastic net solves the following problem

$$\min_{(\beta_0,\beta)\in\mathbb{R}^{p+1}} \left[\frac{1}{2N} \sum_{i=1}^{N} \left(y_i - \beta_0 - x_i^T\beta\right)^2 + \lambda P_\alpha(\beta) \right]$$

where

$$\begin{aligned} P_\alpha(\beta) &= (1-\alpha)\frac{1}{2}\|\beta\|_{\ell_2}^2 + \alpha\|\beta\|_{\ell_1} \\ &= \sum_{j=1}^{p} \left[\frac{1}{2}(1-\alpha)\beta_j^2 + \alpha\,|\beta_j| \right] \end{aligned}$$

is the elastic-net penalty[9]. P_α is a compromise between the ridge-regression penalty ($\alpha = 0$) and the lasso penalty ($\alpha = 1$).

2.2.3 SVM

Support-vector machines is one of the most robust prediction methods for classification and regression. SVMs can efficiently perform a non-linear classification with kernel trick. As the visualization shows, our data is not linear and not very well-separated by label of trait 130. So here we adopt SVMs with radial basis function kernel

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\gamma \left\|\mathbf{x} - \mathbf{x}'\right\|^2\right)$$

2.2.4 Decision Tree

Decision tree learning is a method commonly used in data mining. The goal is to create a model that predicts the value of a target variable based on several input variables. There are two types of decision tree: classification tree and regression tree.

In classification tree, each element of the domain of the classification is called a class. A classification tree is a tree in which each internal (non-leaf) node is labeled with an input feature (Here in our model they are PCA features). The arcs coming from a node labeled with an input feature are labeled with each of the possible values of the target feature or the arc leads to a subordinate decision node on a different input feature. Each leaf of the tree is labeled with a class or a probability distribution over the classes, signifying that the data set has been classified by the tree into either a specific class, or into a particular probability distribution.A tree is built by splitting the source set, constituting the root node of the tree, into subsets—which constitute the successor children. The splitting is based on a set of splitting rules based on classification features.This process is repeated on each derived subset in a recursive manner called recursive partitioning. The recursion is completed when the subset at a node has all the same values of the target variable, or when splitting no longer adds value to the predictions. This process of top-down induction of decision trees (TDIDT)[10] is an example of a greedy algorithm, and it is by far the most common strategy for learning decision trees from data.

In regression tree, it is build up using the same way as classification tree. The only difference is that for each node in regression tree, it is a numeric value instead of a class. It is also learnt by TDIDT.

2.2.5 Random Forest

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean/average prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of overfitting to their training set. The algorithm is comprised of four components: decision tree learning, bagging, from bagging to random forest and extra trees.

The training algorithm for random forests applies the general technique of bootstrap aggregating, or bagging, to tree learners. Given a training set $X = x_1, x_2, ...x_n$ with responses Y $Y = y_1, y_2, ...y_n$, bagging repeatedly (A times) selects a random sample with replacement of the training set and fits trees to these samples:

For a = 1, ..., A:

1. Sample, with replacement, n training examples from X, Y; call these X_a, Y_a.

2. Train a classification or regression tree f_a on X_a, Y_a. After training, predictions for unseen samples x' can be made by averaging the predictions from all the individual regression trees on x':

$$\hat{f} = \frac{1}{A} \sum_{A}^{a=1} f_a(x')$$

3 Results

3.1 EDA

The dataset is imbalanced for Trait 130 (major depression), which is a binary variable here. There are only 95 subjects with major depression. we include all of them as the group of depression. Then, we randomly select 95 subjects from the rest of samples as the no-depression group.

The brain is defined to Brain Regions of Interest (ROIs) corresponding to the nodes in the structural connectivity network based on the popular Desikan-Killiany atlas[3]. The Desikan-Killiany parcellation has 68 cortical surface regions with 34 nodes in each hemisphere which are further divided into 6 left (l) or right (r) domains (FL: Frontal lobe, SU: Insula, LC: Cingulate, PL: Parietal lobe, TL: Temporal lobe, OL: Occipital lobe).

Circular plot of mean of count After calculating the group mean for each connection in these two groups, we display the highly connected edges (thresholded) using circular plots (Figure 2). Overall, the plots from no-depression and depression group look quite similar, but subtle changes can still be seen. For example, there is a connection between ROI3 and ROI17 at the rFL (right frontal lobe) region of the non-depression circular plot which is absent from the circular plot of depression group.

Circular plot of network difference To identify the network difference between no-depression and depression group, we calculate the difference between the mean of each connection in the two groups. The circular plot in Figure 3 shows the top differentiated connections (thresholded). These connections are located in both intra-hemisphere and inter-hemisphere regions. We see decreased connections between lFL and lTL, lPL and rPL, enhanced connections between lFL and rFL, rFL and rLC etc.

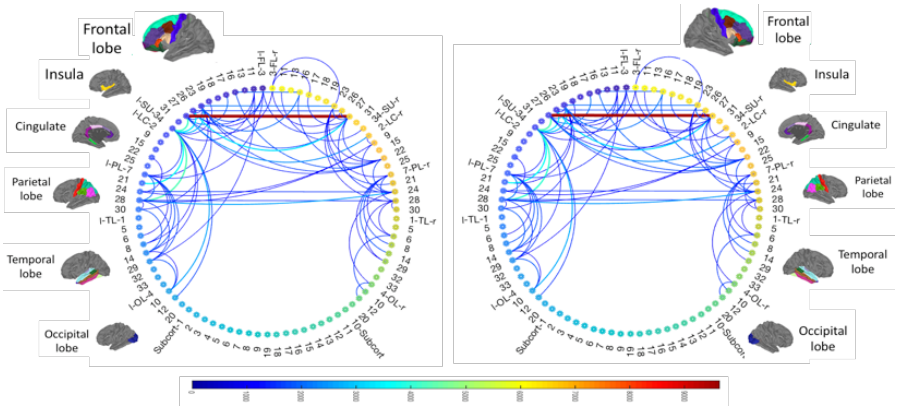


Figure 2: Circular plot of mean of count in the no-depression and depression group. Left plot is for no-depression group and right plot is for depression group. FL: Frontal lobe, SU: Insula, LC: Cingulate, PL: Parietal lobe, TL: Temporal lobe, OL: Occipital lobe, l: left, r: right.

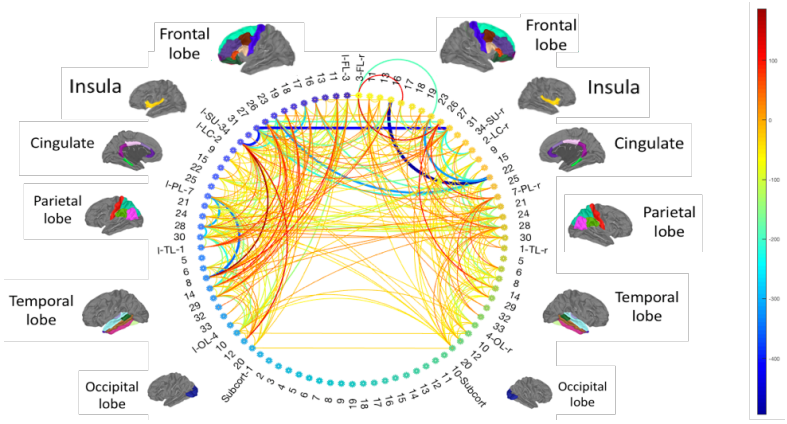


Figure 3: Circular plot of network difference between the no-depression and depression group. FL: Frontal lobe, SU: Insula, LC: Cingulate, PL: Parietal lobe, TL: Temporal lobe, OL: Occipital lobe, l: left, r: right.

For TNPCA scores We also plot the first three PC scores for both FC and SC (Figure 4). The major depression group contains all the 95 subjects who have experienced diagnosed major depression over their lifetime in the dataset. The healthy group also contains 95 subjects, who are randomly selected from the rest.

Figure 4: 3D plots of the first three PC scores for groups with and without major depression experience.

Here we cannot tell a very clear difference in the PC scores distribution of the two groups.

3.2 Hypothesis Testing

Using Connectivity Matrices Next, we perform hypothesis test for each connection between the no-depression and depression group using t-test. The null hypothesis is the connections are similar between the two groups against the alternative they are different. At the cutoff of p-value 0.05, there are 120 edges left. Figure 5 reveals where these 120 connections are located in the brain. We see changes of connections between rFL and rTL, rFL and lLC, rFL and lTL, lTL and lOL, etc. At a more stringent cutoff of p-value 0.01, there are 29 edges left. The circular plot in Figure 6 shows the locations of these 29 connections. The detailed characterization of these 29 connections are listed in Table 1. The observed changes of connections are mainly in the rFL, rTL, lFL, lPL and lTL regions which are consistent with the literature[7]. Interestingly, 3 out the top 5 differential edges are connections inside lPL domain (highlighted in red in Table 1). This PL (Parietal lobe) domain has been reported to be involved in the organization, decision making and predictions of rewards during uncertain conditions[7]. The superior Parietal lobe is also part of the DMN (default-mode network) which is playing a significant role in regulating depression[2]. The significance of these top differential connections awaits further validation and characterization.

ROIa	ROIb	p-value	connection
ROI24	ROI30	0.000194061	lPL-lPL
ROI8	ROI10	0.000458906	lTL-lOL
ROI28	ROI30	0.000623486	lPL-lPL
ROI48	ROI60	0.001086039	rTL-rFL
ROI21	ROI24	0.001206578	lPL-lPL
ROI63	ROI67	0.001319914	rTL-rTL
ROI25	ROI45	0.001472606	lLC-rFL
ROI37	ROI57	0.001827448	rFL-rFL
ROI58	ROI63	0.002164925	rPL-rTL
ROI29	ROI47	0.00220519	lTL-rFL
ROI60	ROI66	0.002507231	rFL-rTL
ROI38	ROI48	0.002840816	rOL-rTL
ROI12	ROI26	0.00303517	lOL-lFL
ROI11	ROI25	0.003421484	lFL-lLC
ROI8	ROI21	0.003967278	lTL-lPL
ROI29	ROI45	0.00444006	lTL-rFL
ROI34	ROI45	0.00444164	lSU-rFL
ROI3	ROI60	0.004506861	lFL-rFL
ROI38	ROI42	0.005439205	rOL-rTL
ROI14	ROI65	0.00556213	lTL-rFL
ROI10	ROI14	0.005865784	lOL-lTL
ROI59	ROI62	0.006738587	rLC-rPL
ROI14	ROI47	0.007244147	lTL-rFL
ROI38	ROI47	0.00825544	rOL-rFL
ROI27	ROI30	0.008485067	lFL-lPL
ROI17	ROI29	0.009026842	lFL-lTL
ROI14	ROI34	0.009645166	lTL-lSU
ROI17	ROI60	0.009777135	lFL-rFL
ROI51	ROI56	0.00990109	rFL-rLC

Table 1: Top 29 differentiated edges (p-value ≤ 0.01). FL: Frontal lobe, SU: Insula, LC: Cingulate, PL: Parietal lobe, TL: Temporal lobe, OL: Occipital lobe, l: left, r: right.

Using TNCPCA scores Here, we use the major depression group and healthy group as mentioned above. We first do a simple two-sample t test to test the mean difference between their first PC score. The p-value is 0.88 for FC and 0.29 for SC. Neither of the tests reject the null hypothesis that the mean of the first PC scores are different for the two groups. Then, we apply the maximum mean discrepancy (MMD)[4] to test whether the two samples are drawn from different distributions using all the 60 PC scores. The test results suggest that, for both FC and SC, we do not have enough evidence to say that the brain PC scores distributions are different for the two groups.

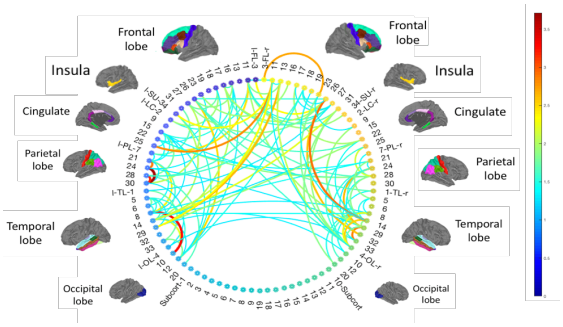


Figure 5: Circular plot of top differentiated edges (p-value ≤ 0.05, edge number = 120) between the no-depression and depression group. The plot was made using -log10 p-values. The connections inside lPL domain are colored as red. Frontal lobe, SU: Insula, LC: Cingulate, PL: Parietal lobe, TL: Temporal lobe, OL: Occipital lobe, l: left, r: right.

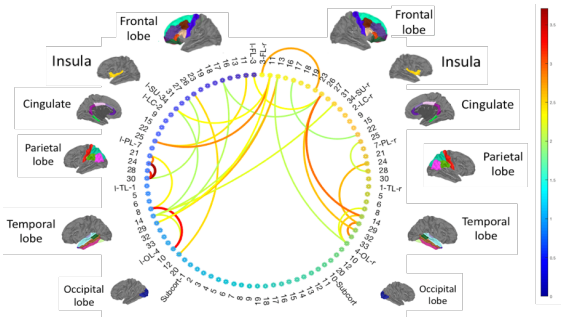


Figure 6: Circular plot of top differentiated edges (p-value ≤ 0.01, edge number = 29) between the no-depression and depression group. The plot was made using -log10 p-values. FL: Frontal lobe, SU: Insula, LC: Cingulate, PL: Parietal lobe, TL: Temporal lobe, OL: Occipital lobe, l: left, r: right.

3.3 Modelling Results

We are interested in studying the relationship between depression and brain connectome for all the HCP subjects. Particularly, we want to see whether depression can be predicted based on brain connectomes. Here, we try to predict depression using multiple machine learning methods with human connectome data–i.e., SC

and FC TN-PCA data. We randomly select 1/3 data as test set and use the rest data as training set. As mentioned earlier, we also add 200 synthetic data to the training set to overcome the imbalance of the original data set. In this section we are going to show the modelling results for both Trait 130: Major Depression Episode and Trait 131: Number of Depression Symptoms. We will introduce the result for each Trait seperately.

3.3.1 Classification: Major Depression Episode

Logistic Regression, LDA and Naive Bayes Table 2 summarizes the results of these three methods. Because of the imbalance of the dataset for Trait 130, although all three methods show relative high accuracy, the sensitivity is low. Logistic Regression and LDA don't perform well here, even slightly below the chance level based on the values of Area Under Curve (AUC, Table 2). In contrast, Naïve Bayes model gives a better than random performance (classification rate: 72.8%, sensitivity: 27.3%, specificity: 77.8% and AUC score: 0.5579, Table 2 and Figure 7).

	Accuracy	Specificity	Sensitivity	AUC
Logistic Regression	80.30%	86.75%	21.21%	0.4692
LDA	81.19%	88.41%	15.15%	0.4714
Naïve Bayes	72.84%	77.81%	27.27%	0.5579

Table 2: Performance Measures of Logistic Regression, LDA and Naive Bayes models

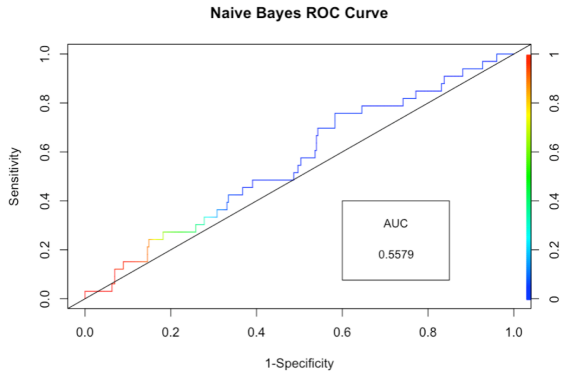


Figure 7: ROC Curve of Naive Bayes Model

Ridge Regression, Lasso and Elastic Net Since ridge regression, lasso and elastic net give the fitted probabilities of binomial regression, we first choose a threshold to make the decision. Figure 8 shows the relationships between the threshold and the sensitivity or specificity. For example, for ridge regression with only SC features as predictors, if we pick a threshold of 0.4, we get the following confusion matrix.

Prediction \ Reference	0	1
0	259	24
1	43	9

Table 3: Confusion Matrix for ridge regression with SC features

Here we try to include only SC features, only FC features and all the features into models and seems SC is more reliable for all of this methods. So these three models are based on the 60 SC features. Figure 9 shows the ROC for ridge regression, lasso and elastic net with alpha = 0.8. We can see that they all have a AUC a little bit higher than 0.5. Table 4 is the comparison on the 3 models, there is no significant difference. they all have accuracy over 70% but have a low sensitivity.

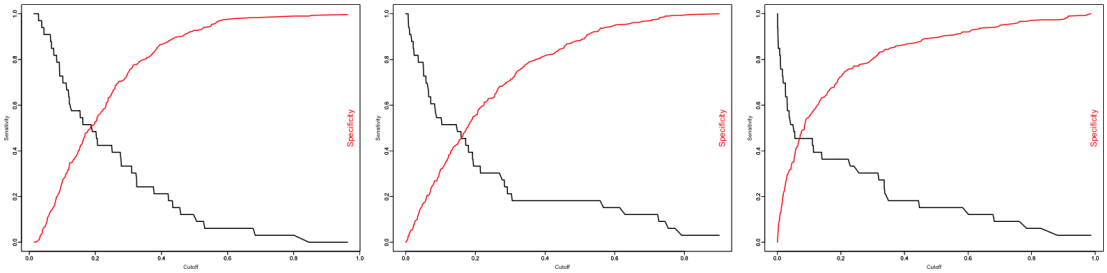


Figure 8: Cutoff v.s. Sensitivity or Specificity for lasso models with only FC features, SC features and all the features.

	AUC	Accuracy	Specificity	Sensitivity
alpha = 0	0.5123	80 %	85.76 %	27.27 %
alpha = 1	0.5131	74.63 %	80.13 %	24.24 %
alpha = 0.8	0.5128	80.9 %	87.42 %	21.21 %

Table 4: Confusion Matrix for ridge regression with SC features

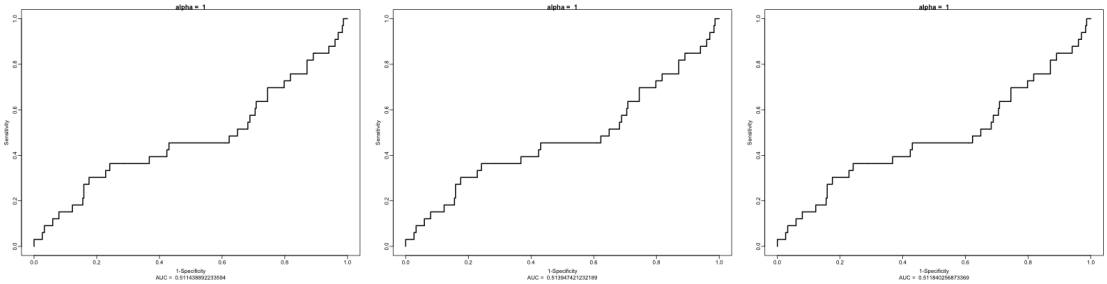


Figure 9: ROC for rigde regression(alpha = 0), lasso(alpha = 1) and elastic net(alpha = 0.8).

SVM For SVM, we also try many combinations of predictors. For instance, if we throw all the 120 features in, the model will never give a positive prediction. It is the same problem as before, the data is not well separated by the label. Here is the result for the SVM with top 10 FC features and top 10 SC. We have an AUC of 0.53 and the accuracy is 86%, but still we have a low true positive rate.

Prediction \ Reference	0	1
0	286	31
1	16	2

Table 5: Confusion Matrix for SVM

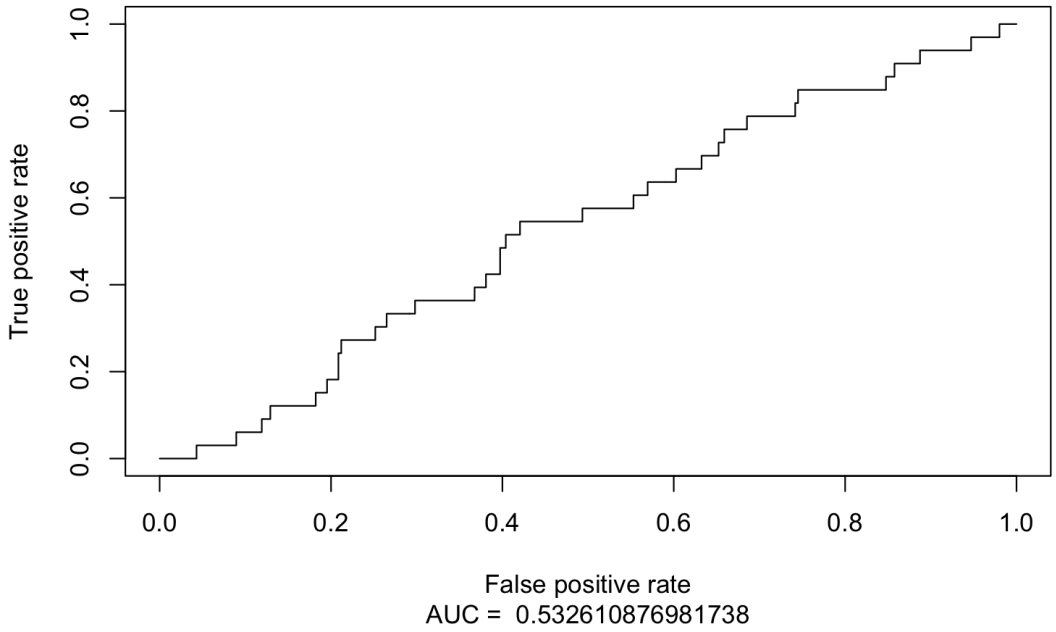
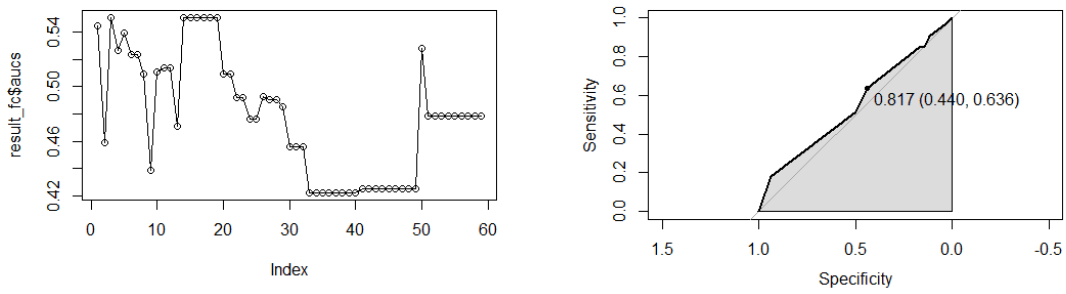


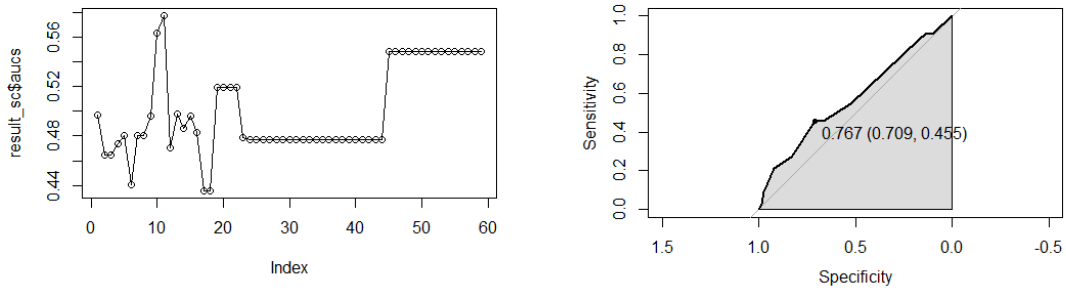
Figure 10: ROC for SVM.

Decision Tree For decision tree, we tried three experiments: FC features, SC features and the combination of FC features and SC features. We use cross validation to train the model. Here we need to determine how many features as input to train the tree model could achieve the best performance. So we experiment features from 0 to all for each feature set. Here we use AUC as the evaluation metric. The change of AUC with the number of features and the ROC curve under the best performances for each feature set is shown as Figure 11

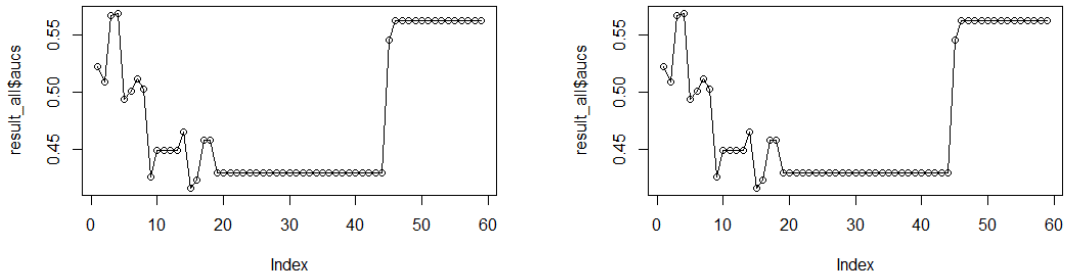
We can see that for FC, SC and the combination of FC and SC, the best



(a) FC features



(b) SC features



(c) FC and SC features

Figure 11: Change of AUC with feature numbers and ROC of the best performance for different combinations of features

performance is achieved when the features are round 15, 11 and more than 45 separately.We can also see that when using the combinations of features, it needs more features to achieve the best performance. The summary of the best performance for each feature set is shown in Table 6. The best AUC is achieved when only using SC features while the worst is achieved when only using FC features.

	FC	SC	FC+SC
AUC	0.550	0.578	0.569
Accuracy	0.460	0.684	0.343
Threshold	0.817	0.767	0.833

Table 6: AUC, Accuracy and Threshold for different combinations of feature sets for decision tree

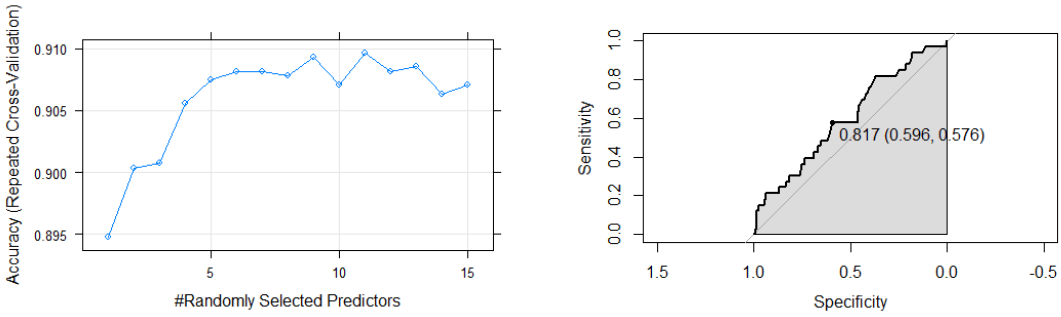
Random Forest For random forest, we do the same three experiments as the decision tree. We also use cross valiation to train the model and this time we need to decide the number of mtry. For each combination of feature set, we try from 1 to 15 as the value of mtry. Here we still use AUC as the metric. The change of the performance of it with number of mtry for each combination of feature set and the ROC curve under the best performance is shown in Figure 12.

We can see that for FC, SC and the combination of FC and SC, the best performance is achieved when mtry is around 10. The summary of accuracy, AUC and threshold is shown in Table 7.

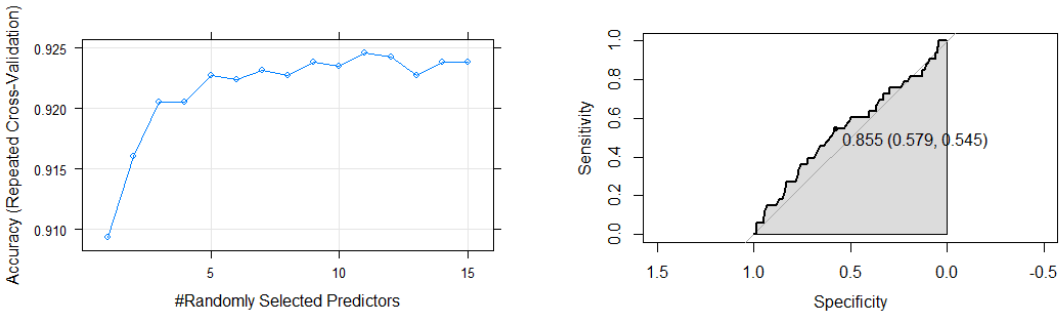
	FC	SC	FC+SC
AUC	0.600	0.549	0.551
Accuracy	0.594	0.576	0.573
Threshold	0.817	0.855	0.829

Table 7: AUC, Accuracy and Threshold for different combinations of feature sets for random forest

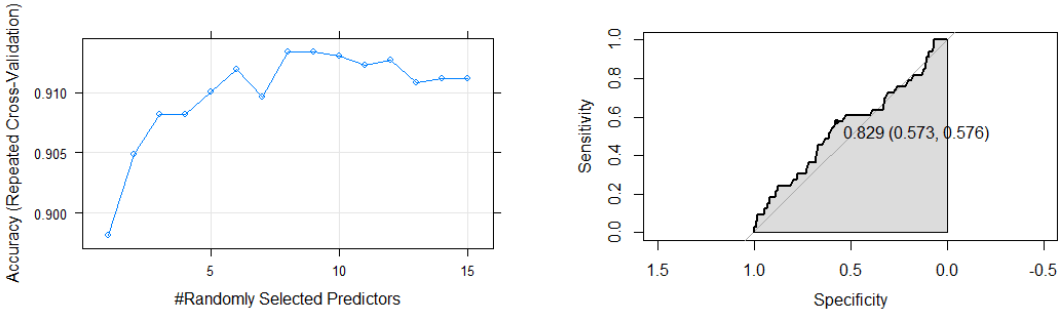
We can see the best performance is achieved when using FC feature only and the AUC is 0.600 while it is 0.578 for a single decision tree. Random forest achieves better result than a single decision tree for this task.



(a) FC features



(b) SC features



(c) FC and SC features

Figure 12: Change of AUC with feature numbers and ROC of the best performance for different combinations of features for decision tree

3.3.2 Regression: Number of Depression Symptoms

Trait 131 illustrates the number of depression symptoms, which is a continuous data. Here we fit a linear regression as the baseline to compare the MSE for all methods. The results are shown in Table 8. As we can see, the simple models like linear regression and decision tree have lower MSE. One reason can be that the data is very unbalanced such that a simple model is preferred. The histogram is shown in Figure 13. So it is not surprising that ridge regression, lasso, SVM and random forest give even worse results than the linear regression.

Since we have a very large proportion 0's in trait 131, we fit new models with all 0's removed. The results in Table 9 shows a improvement of model fitting then that with unbalanced data. Now, the simple model like decision tree does not perform as well as before. Except ridge regression, all other models have better prediction than the baseline.

	FC	SC	FC+SC
linear regression (baseline)	6.64	6.67	7.06
ridge regression	7.99	6.20	7.66
lasso	6.27	6.41	6.48
SVM	8.75	7.98	7.88
Decision tree	1.57	1.80	1.90
Random forest	7.58	7.03	6.52

Table 8: MSE for regression models on trait 131

	FC	SC	FC+SC
linear regression (baseline)	4.48	8.29	16.07
ridge regression	4.97	4.94	5.33
lasso	3.63	3.55	3.57
SVM	4.29	4.78	4.54
Decision tree	5.77	7.09	5.49
Random forest	3.69	3.56	3.44

Table 9: MSE for regression models on trait 131 with 0's removed

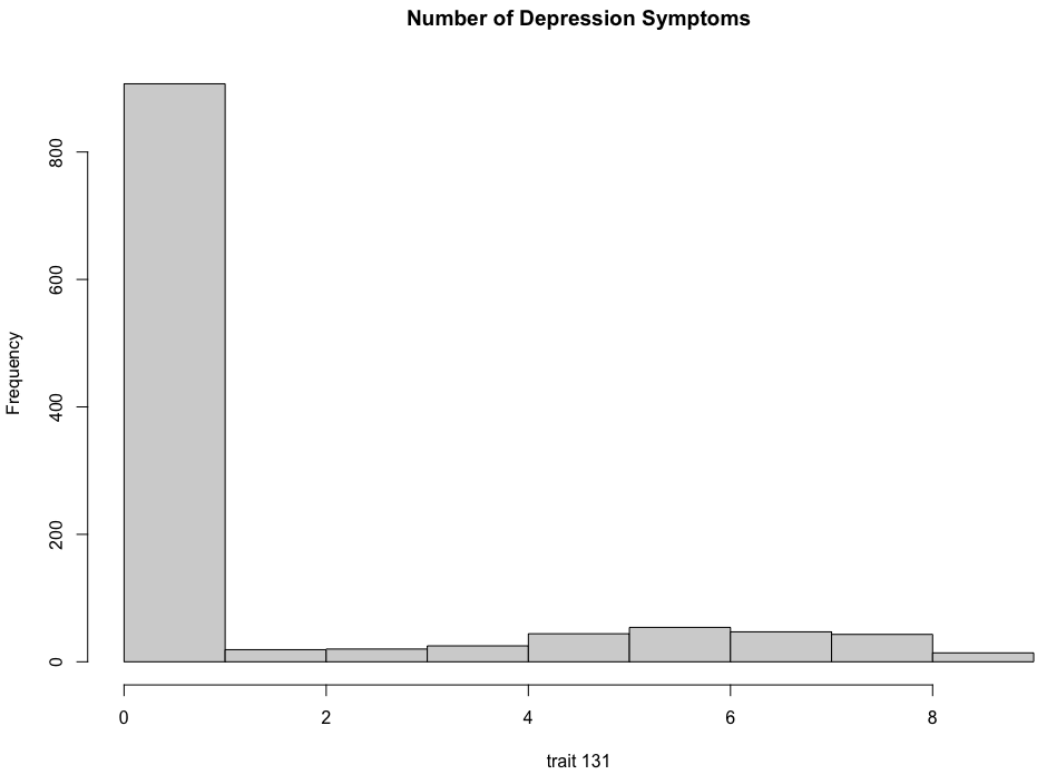


Figure 13: Distribution of Trait 131: number of depression symptoms.

4 Discussion

In our study, we try to study the relationship between HCP connectome and depression in a cohort of about 1,000 adults. Through hypothesis testing, we don't observe significant differences in the distribution of the brain PC scores among subjects with a low versus high depression score. However, we do find some potentially interesting alterations of connections in the rFL, rTL, lFL, lPL and lTL regions which were previously reported to be involved in the regulation of depression[7]. The significance of the changes of these connections needs further validation and characterization in the future.

We also try to predict depression based on the connectome data using multiple

state-of-art machine learning approaches. The Naive Bayes, Penalized Regression (Ridge, Lasso and Elastic Net) and SVM give slightly better than chance performance, which are in line with a couple of recent studies including a similar large-scale study done in 10,343 healthy individuals from UK Biobank[5].

There is a lack of consistency and generalizability among the published literature of relationship between brain connectome and depression. The reasons are likely, at least partly, due to the differences in symptom profiles, sample characteristics, dissimilar definitions of ROIs, different methods for measurement and power of analytical tools[7]. However, Anterior Cingulate Cortex (ACC), Hippocampus and Amygadala have been consistently shown to be important in the regulation of depression. Hippocampus and Amygadala are located in the subcortical region, which are missed in our study because of the parcellation method. But we do detect changes of connections between ACC and other domains (such as Frontal lobe and Parietal lobe, Table 1).

Therefore, with the improvement of measurement and analytical tools, dissecting the relationship between the brain structural and functional connectome and depression, in the long run, can help us better known the pathology and will promote prevention, early diagnosis and treatment of major depressive disorder.

5 CRediT authorship contribution statement

- **Ling Cai**: Formal Analysis (Hypothesis Testing, Logistic Regression, LDA and Naive Bayes), Resources, Software, Visualization, Writing (Discussion).
- **Zhaoqi Liu**: Data Curation, Formal Analysis (Hypothesis Testing), Project administration, Resources, Software, Visualization, Writing (Introduction).
- **Ximing Wen**: Formal Analysis (Regression Tree, Random Forest), Software, Visualization, Writing.
- **Wan Zhang**: Data Curation, Formal Analysis (Penalized Regression, SVM), Software, Visualization, Writing.

References

[1] Contreras JA, Goñi J, Risacher SL, Sporns O, Saykin AJ. The Structural and Functional Connectome and Prediction of Risk for Cognitive Impairment in Older Adults. Curr Behav Neurosci Rep. 2015;2(4):234-245. doi:10.1007/s40473-015-0056-z

[2] De Witte NAJ, Mueller SC. White matter integrity in brain networks relevant to anxiety and depression: evidence from the human connectome project dataset. Brain Imaging Behav. 2017 Dec;11(6):1604-1615. doi:10.1007/s11682-016-9642-2. PMID: 27744495.

[3] Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage. 2006 Jul 1;31(3):968-80. doi: 10.1016/j.neuroimage.2006.01.021. Epub 2006 Mar 10. PMID: 16530430.

[4] Gretton A., Borgwardt KM., Rasch MJ, Schölkopf, B, Smola A. A kernel two sample test. 2012. J. Mach. Learn. Res. 13, 723–773

[5] Maglanoc LA, Kaufmann T, van der Meer D, Marquand AF, Wolfers T, Jonassen R, Hilland E, Andreassen OA, Landrø NI, Westlye LT. Brain Connectome Mapping of Complex Human Traits and Their Polygenic Architecture Using Machine Learning. Biol Psychiatry. 2020 Apr 15;87(8):717-726. doi: 10.1016/j.biopsych.2019.10.011. Epub 2019 Oct 29. PMID: 31858985.

[6] Pandya M, Altinay M, Malone DA Jr, Anand A. Where in the brain is depression?. Curr Psychiatry Rep. 2012;14(6):634-642. doi:10.1007/s11920-012-0322-7

[7] Zhang FF, Peng W, Sweeney JA, Jia ZY, Gong QY. Brain structure alterations in depression: Psychoradiological evidence. CNS Neurosci Ther. 2018;24(11):994-1003. doi:10.1111/cns.12835

[8] Zhang Z, Allen GI, Zhu H, Dunson D. Tensor network factorizations: Relationships between brain structural connectomes and traits. Neuroimage. 2019;197:330-343. doi:10.1016/j.neuroimage.2019.04.027

[9] Zou H, Hastie T. Regularization and variable selection via the elastic net. J. Royal. Stat. Soc. B., 67(2):301–320, 2005.

[10] Quinlan, J. R.Induction of decision trees. Machine Learning. 1: 81–106, 1986