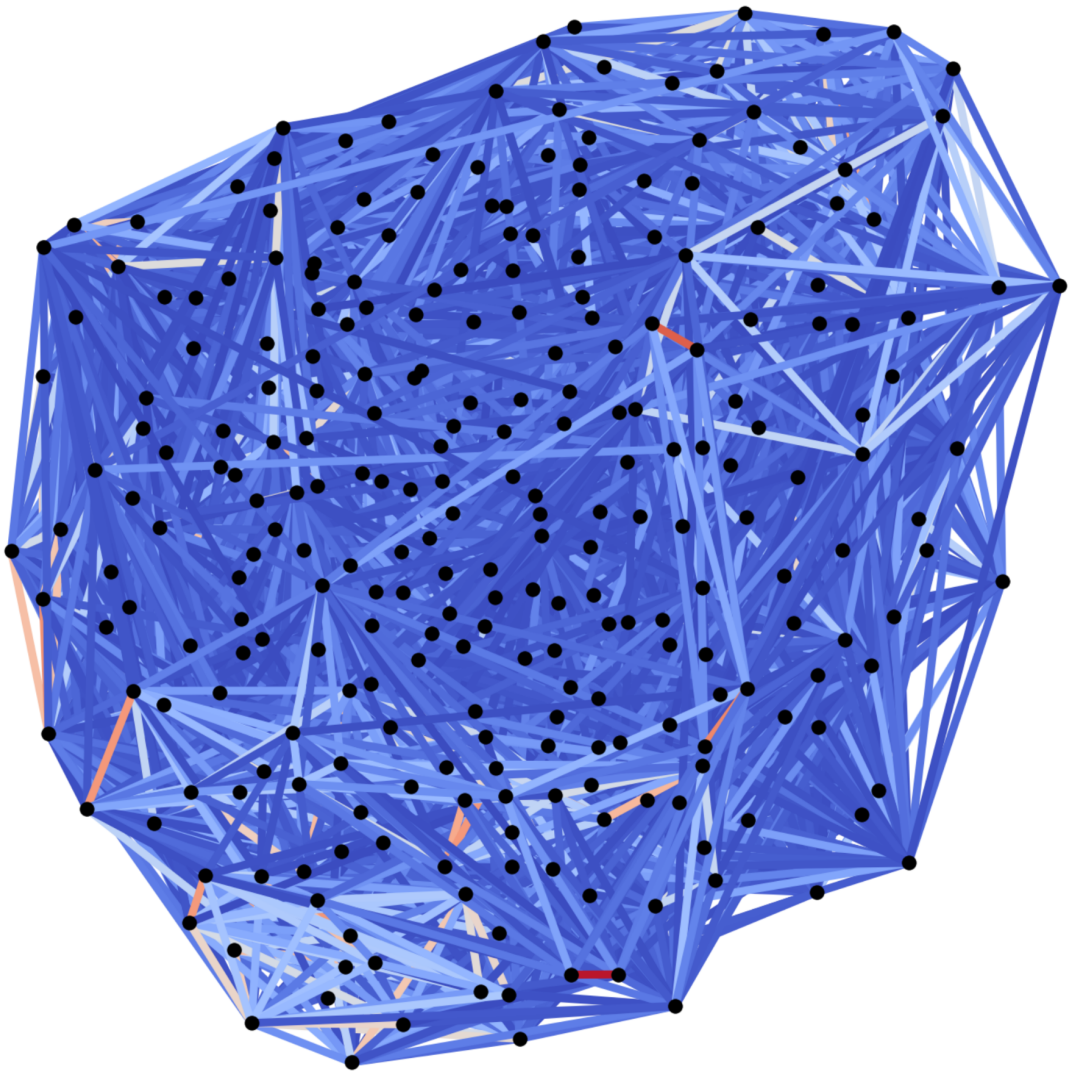
rs-fMRI Graph Theoretical Exploration of Traumatic Brain Injury Subjects

CSB199 Report

Daniel Frees
UID: 605104056
03/17/2022



Img: Brain Graph for Subject 1 (Spring Layout)

Abstract

This project was designed and undertaken with the goal of determining potentially significant differences in graph theory measures of injured brains as compared with healthy brains, as well as to determine whether graph theoretical properties derived from resting state functional MRI (rsfMRI) images might provide valuable insights into the underlying function of traumamtic brain injury (TBI) in the brain.

To accomplish this goal, the first step was extracting weighted adjacency matrices from rsfMRI images. BrainSport's CBF dataset provided 104 participants with complete imaging sessions, and CONN Toolbox was used to preprocess these images, run connectivity analysis, and output connectivity matrices with edge weights corresponding to Fisher's Z values between ROIs. In addition to connectivity matrices, relevant covariates (including the target: presence of traumatic brain injury) were extracted. Data was cleaned, imputed, thresholded, and wrangled through MATLAB and Python, primarily using Python's pandas package. Once the data was cleaned, Python's NetworkX package and custom scripts were used to calculate graph theory measures for each subject in the dataset, with a focus on global metrics and amygdala region of interest (ROI) metrics. Then, the means and variances of these metrics were compared between mild TBI (mTBI) and HC groups, and t-tests were used to determine significantly different metrics. Numerous machine learning algorithms were explored with the goal of developing a model which could accurately predict whether a subject had a brain injury based on their rsfMRI graph metrics. Hyperparameters were tuned on six model archetypes, and the best validated models' were evaluated based on their performance on unseen testing data.

Several findings bordering on significance came out of this exploratory analysis. Chief among these was the finding that transitivity may be different in TBI brains compared to healthy control brains (2-sided $p<0.075$). Transitivity was found to be moderately negatively correlated with TBI, suggesting that brain cliques might break down with injury.

Unfortunately, none of the machine learning models trained to predict TBI compared to HC were able to perform significantly better than a naive majority vote prediction accuracy, indicating that the provided rsfMRI graph metrics are not strong predictors of TBI. It should further be noted that there was likely not enough data to properly learn the complex functionality underlying TBI. It is possible that with a much larger dataset, a better ML model could be developed for the purpose of predicting presence of injury based on rsfMRI graph theory metrics.

Introduction

Resting-state functional MRI (rsfMRI) is currently under research for its potential uses for diagnosing TBI and predicting symptom outcomes following insult. Images are taken at a resting-state, meaning that subjects perform no specific tasks while undergoing imaging, and

are usually told to think about nothing in particular. rsfMRI measures blood-oxygen-level-dependent (BOLD) signal throughout the brain, and can be used to compute correlations between brain regions under the assumption that BOLD signals correlate over time between regions that are more highly correlated (brain regions with BOLD signals which anticorrelate are considered to be negatively correlated).

Despite the resting-state nature of rsfMRI, controlling covariates to easily differentiate between brain injury patients and healthy control subjects remains a challenge. Differences in brain shape, movement, thought processes, personality, gender, handedness, lesion status, injury severity, chronicity (time since injury), and other variables can all affect rsfMRI results. As such, rsfMRI studies have found inconsistent trends when investigating specific brain networks/ regions. For this reason, graph theoretical measures which summarize larger scale patterns in the brain might be a valuable tool for investigating rsfMRI results.

Studies aiming to diagnose autism using machine learning models have found greater success when incorporating topological graph theory metrics from patient resting-state functional MRI images[1]. The goal of this study is to extend a similar approach to the classification of traumatic brain injury. In this report, an exploratory data analysis is undertaken to identify graph theoretical properties which might serve as useful features to input into an ML classification algorithm, as well as to determine significant differences between mTBI and HC brains using classical statistical methods. Furthermore, numerous ML models are explored to predict presence of TBI, and their pitfalls are described.

Methods

Extraction: Using MATLAB's CONN Toolbox[2], images from time point 1 of BrainSports' CBF dataset were preprocessed with standard regressions, smoothing, and motion correction. Regions of interest (ROIs) were selected based on the Brainnetome 264 Atlas[3]. Analyses were performed to calculate Fisher's z value correlations between all pairs of ROIs for each subject, and this data was exported in .mat files containing adjacency matrices and numerous other results. Covariates were also exported in an .xlsx file.

MATLAB was used to extract adjacency matrices from amongst the large variety of data contained in the .mat results output file. Next, all data was imported into a Python Jupyter Notebook and imported as pandas dataframes. Data was cleaned, imputed, thresholded, and wrangled primarily using Python's pandas package. Once the data was cleaned, Python's NetworkX package and custom scripts were used to calculate graph theory measures for each subject in the dataset, with a focus on global metrics and amygdala ROI metrics. The following analysis steps are unique to this project, but based largely on prior precedents outlined in the primary paper establishing BRAPH (Brain analysis using graph theory) software[4].

Graph Theory Metric Calculation & Graph Theory Based Processing:

1. ROI-ROI correlations zeroed for ROIs to themselves (zero diagonal of adj. mat.)
2. Negative correlations zeroed
3. All subject adjacency matrices passed through a dynamic correlation threshold designed to make all graphs have density ~ 0.25
4. ROIs of interest (nodes_of_interest) extracted from the entire list of ROIs. For this, all amygdala ROIs were extracted as ROIs of interest.

For each subject:

5. Covariate data matched to subject
6. NetworkX graph generated from subject adjacency matrix
7. Weighted and unweighted clustering coefficients calculated
8. Characteristic path length calculated
9. Density calculated to verify that all patients have density around 0.25 (minor fluctuations may occur due to limitations of float precision and ties around the dynamic threshold)
10. Transitivity calculated
11. Node degree and node strength calculated for all ROIs of interest
12. Closeness centrality calculated for all ROIs of interest
13. Shortest path calculated between left dorsal amygdala ROI and left MB ROI
14. Shortest path calculated between left dorsal amygdala ROI and right MB ROI

Statistics and Visualization:

Upon completion of graph theory processing and metric calculations, histograms were visualized for all variables in the dataset, as well as separate histograms for the TBI data subset and HC data subset. Means and standard deviations were calculated for each metric within the TBI and HC subsets. Two-sample Welch's t-test using two-sided alternative hypotheses were calculated to compare features across the TBI and HC groups. Finally, correlations were visualized for the entire dataset.

Machine Learning Modeling:

First, a widespread exploratory analysis was conducted to identify pitfalls of the dataset. A few of the main problems identified through data manipulation and poor performance on basic model implementations:

- The Brainsport CBF dataset is highly imbalanced, with nearly twice as many healthy controls as TBI patients. This creates problems training models where they may optimize by simply predicting the majority class, and struggle to learn anything valuable.
- The dataset is very small, making undersampling to adjust imbalanced classes problematic, and learning complex functions impossible.
- Traumatic brain injury is not known to be characterized by monotonic differences in brain metrics, but rather is expected to present as a combination of different effects, unique to each individual subpopulation of the injury. This suggests that TBI would be best learned

by a complex feature space separating model such as Random Forests or Adaboost. Unfortunately, to develop complex feature space separations without overfitting training data requires significant amounts of training data and clever regularization techniques.
- Many graph metrics yield relatively similar values for different brains under the preprocessing techniques supported by current brain graph theory research. Noise introduced by imaging abnormalities, movement, phenotypic brain structure differences, patient thoughts, and more could be enough to upset these graph metric values and overshadow a significant difference between groups that theoretically exists.

Next, scripts were developed to deal with as many of the aforementioned issues as possible. Increasing the amount of samples is impossible, but oversampling the existing samples with replacement balances classes so that models can be trained on balanced data (and validation accuracies can be interpreted relative to a naive accuracy of 0.5). A script was developed to balance data by either oversampling or undersampling. Furthermore a split-generating class was written to pass to scikit-learn's GridSearchCV to oversample training and validation data separately and prevent data leakage due to pre-split oversampling.

A feature selection function was developed to pass only features falling under a certain t-test p-value threshold to the ML model testing framework. The goal was to reduce the number of features models had to learn. Originally, features nearly outnumbered the number of samples in the training data, making overfitting and poor testing performance more likely.

A comprehensive ML model testing framework was designed to test six ML model archetypes (K-Nearest Neighbors (KNN), Decision Tree (DT), Support Vector Classifier (SVC), Logistic Regression (LogReg), Gaussian Naive Bayes (Gaussian NB), and Random Forest (RF)). In this framework, hyperparameters were tuned via evaluation on stratified oversampled 10-fold cross-validation generalization accuracy, and the best cross-validated performance model was selected for testing on previously unseen testing data.

Finally, performances of the ML models from the above testing framework were compared and plotted relative to naive testing accuracy.

Results

Distribution of Features (Complete Dataset):

Some interesting distributions appeared when visualizing all features using histograms. First of all, it should be noted that the raw dataset is imbalanced, with approximately ~⅓ of the data corresponding to TBI patients and ~⅔ to HC subjects. This should not affect the current analysis, but will important for future directions with machine learning. Average clustering coefficient was approximately normally distributed for both weighted and unweighted clustering coefficients. Characteristic path length was strongly right skewed. Transitivity did not match any canonical distributions but was approximately right skewed. Node degree for both the left dorsal

amygdala ROI and right dorsal amygdala ROI were bimodally distributed. Closeness centrality for the left lateral amygdala was bimodally distributed as well, as was closeness centrality for the right medial amygdala.

See distributions on the following page.

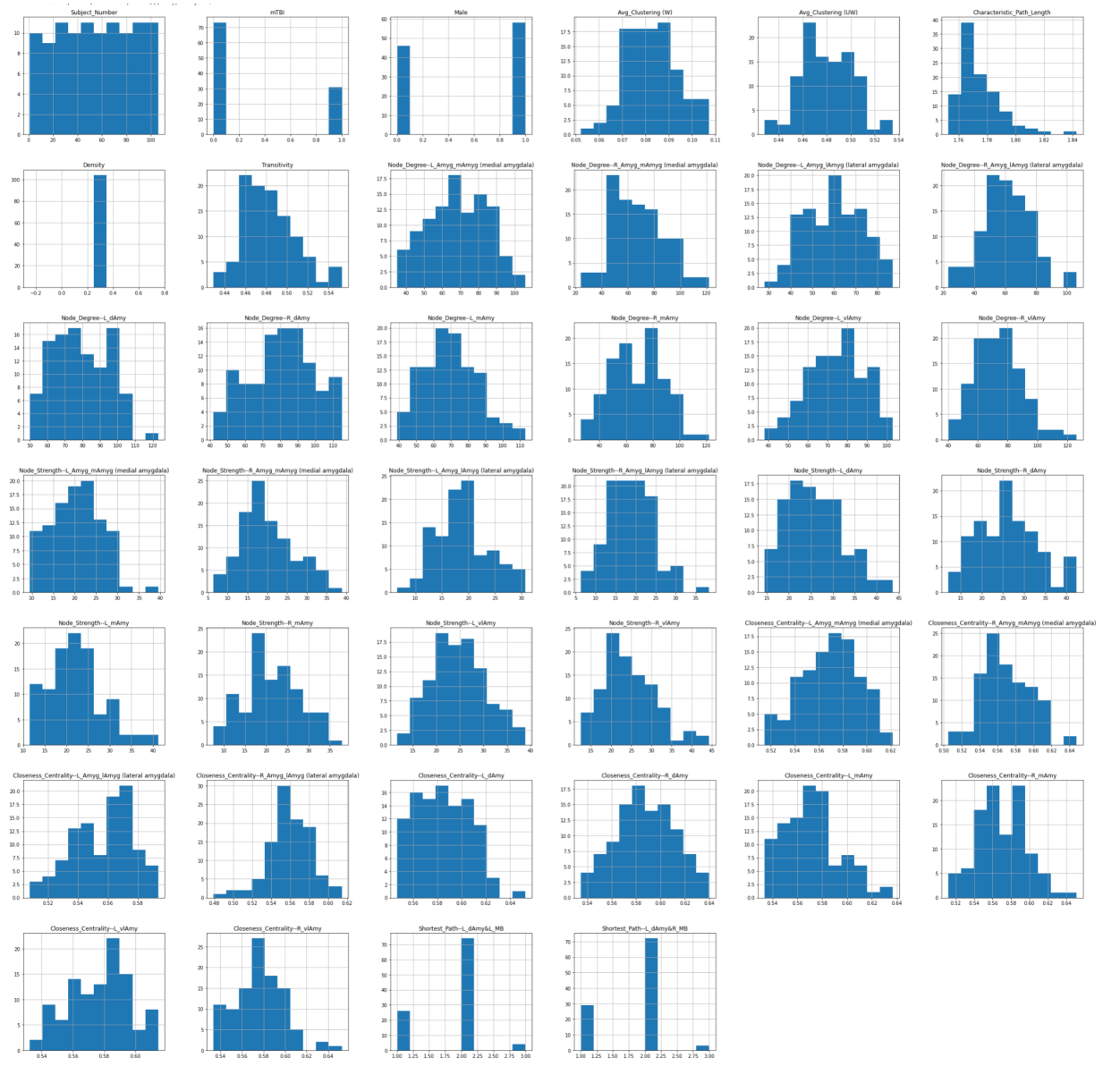


Figure 1: Distributions of Graph Theory Features Across Full Dataset

Correlation Analysis (All Data):

For the following correlation analysis, note that presence of mild TBI (mTBI) was indicated with a 1, and absence was indicated with a 0. Thus, positive correlations between any feature and *mTBI* indicate that said feature trends higher in the *mTBI* population.

mTBI is moderately negatively correlated with average unweighted clustering coefficient, characteristic path length, transitivity, both node degree and node strength for the right dorsal amygdala ROI, closeness centrality for the right dorsal amygdala ROI, and shortest path length between the left dorsal amygdala and left MB.

Furthermore, mTBI is moderately positive correlated with closeness centrality of the left lateral amygdala.

See the following correlation outputs. Note that all values reflect Pearson's r values, and all correlations visualized below are with *mTBI.*

Figure 2: Visualization of Correlations (Pearson's r) of All Features with mTBI

Machine Learning Model Performance

Unfortunately, all models evaluated in this project performed poorly (Figure 3). While several models (KNN, SVC, LogReg) were able to achieve the naive expected performance (defined as the accuracy that a majority vote classifier would output) of 0.65625, none outperformed this naive accuracy. The worst model by far was Gaussian Naive Bayes, with a testing accuracy of 0.46785.

```
Expecting a naive TESTING accuracy of 0.65625
Note: Validation Data is Balanced Separately Among Train/Val split data...
Therefore, validation naive accuracy = 0.5

----------------------KNN-----------------------
Best KNN Model: KNeighborsClassifier(n_neighbors=17)
Best KNN CV Validation Accuracy:0.6799999999999999
Best KNN (Single Run) Testing Accuracy: 0.65625
------------------------------------------------

--------------------DT--------------------------
Basic DT Test Accuracy: 0.5625
Best Decision Tree: DecisionTreeClassifier(max_depth=7, max_features='sqrt', random_state=42)
Validation Accuracy: 0.5933333333333335
Test Accuracy: 0.625
------------------------------------------------

--------------------SVC-------------------------
Basic (RBF, Scale) SVC Accuracy: 0.65625
Best SVC: SVC(degree=1, kernel='poly', shrinking=False)
Best SVC Validation Accuracy: 0.6583333333333333
Best SVC Test Accuracy: 0.65625
------------------------------------------------

-------------------LogReg-----------------------
Basic LogReg Test Accuracy: 0.65625
Best Logistic Regression: LogisticRegression(C=0.5, max_iter=10000, penalty='l1', solver='liblinear')
Best LogReg Validation Accuracy: 0.6966666666666665
Best LogReg Test Accuracy: 0.65625
------------------------------------------------

-----------------Naive Bayes--------------------
Gaussian Naive Bayes Test Accuracy: 0.46875
------------------------------------------------

----------------Random Forest-----------------
Basic RandomForest Test Accuracy: 0.625
Best RF Model: RandomForestClassifier(max_depth=25, max_features='sqrt', random_state=42)
Validation Accuracy: 0.5666666666666667
Test Accuracy: 0.625
------------------------------------------------
```

Figure 3. Poor ML Model Performance Relative to Majority Vote

Below, find a comparison of testing performance among the tuned models for each archetype in bar plot form.

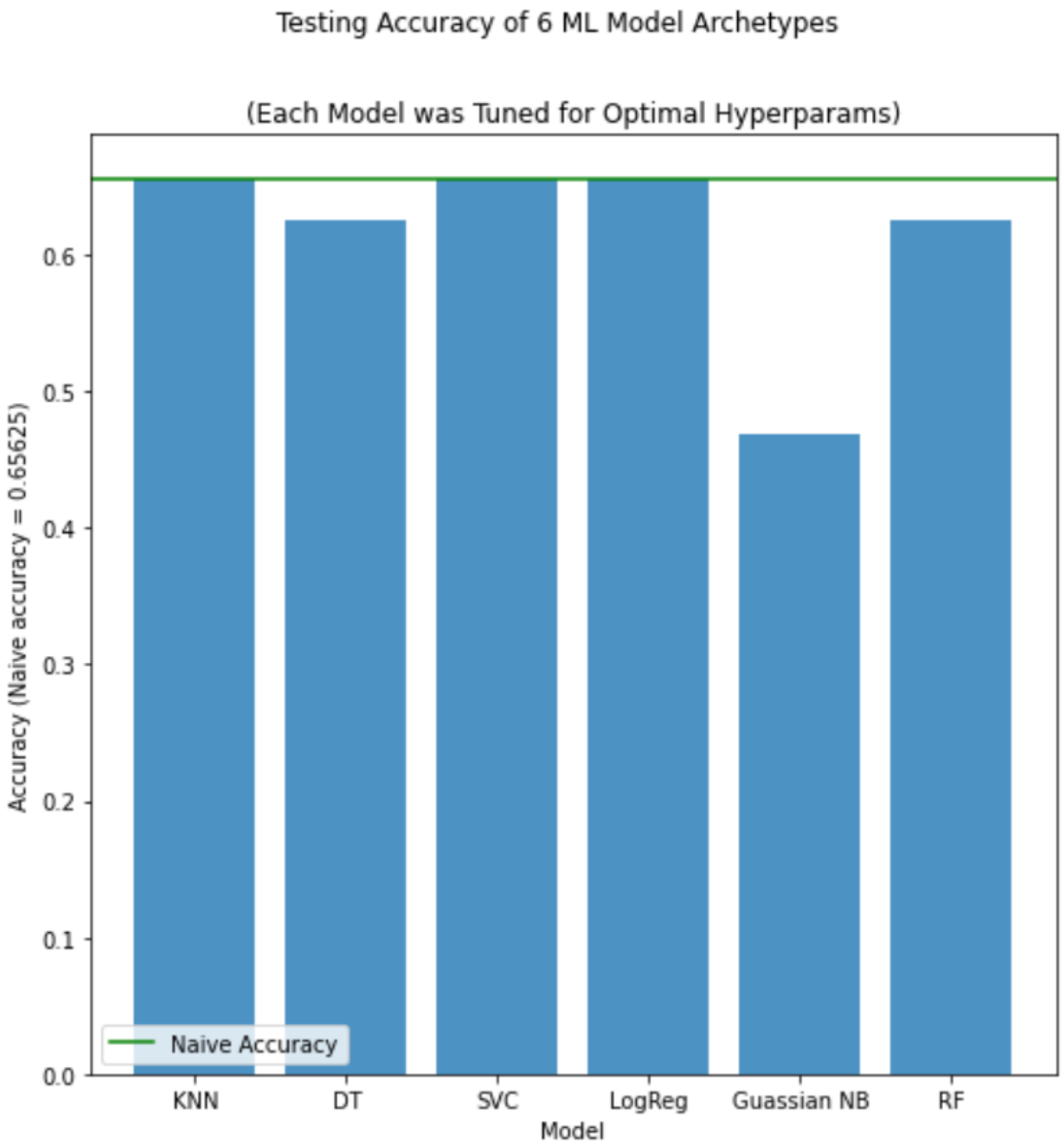


Figure 4. Bar Plot of Model Testing Accuracies

Discussion

Feature Distributions:

The current analysis is not sufficient to explain the underlying causality of these various distributions, however the existence of bimodal distributions is promising. It may be the case that either TBI or some subset of TBI correlates to one of the modes in some of these distributions. Based on the machine learning analysis and t-testing, however, it seems more likely that these bimodal distributions correlate to some other subgrouping of the subjects.

Classical Statistical Analysis:

Subject_Number is significantly different between groups only because subject numbers were ordered from mTBI participants to HC participants, and therefore indicates nothing. *mTBI* is significant because it is the way the groups were formed, and also indicates nothing. *Density* is significant only because of minor differences in node thresholding (tied nodes can ever so slightly affect the manually chosen 25% density conformity requirement imposed in the dynamic thresholding function). There is no difference between groups, but the standard deviation of density is so low that analysis yields a very low p-value.

Transitivity on the other hand, is an interesting group difference finding. While the p-value would only be semi-significant at alpha = 0.05, this is still a promising finding. Transitivity is a measure which can be interpreted in terms of cliquiness of a graph, whereby higher transitivity indicates that node relationships are more likely to form a clique. Specifically, transitivity measures the chance that if node x knows node y (is correlated), and y is correlated with z, then x is also correlated with z. Based on this data, transitivity seems to be higher for the healthy control group (mu=0.4857) than the traumatic brain injury group (mu=0.47704), indicating that perhaps traumatic brain injury breaks down cliques. This is also indicated by the negative correlation between *mTBI* and *Transitivity* in the correlation analysis. Transitivity should certainly be hypothesized to be an important feature for future machine learning modeling.

Correlation Analysis:

Correlation calculation can be misleading, so no conclusions will be drawn firmly from the presented correlation values. However, the negative correlation of TBI with transitivity suggests that cliques may break down in TBI.

ML Modeling

All models performed poorly, even after feature selection, hyperparameter tuning and specialized cross-validation scripts designed to deal with the small unbalanced dataset. This suggests that the provided data was insufficient to learn to predict brain injury. Either more data or different features are required to learn a valuable prediction algorithm.

Specifically, the poor model performances relative even relative to majority vote classification suggests that the models were unable to learn patterns in the data which generalized to the testing data. In some cases, models performed worse than the naive accuracy of 0.65625, which is likely due to overfitting to noise in the training and validation data.

The particularly poor performance from the Gaussian Naive Bayes model suggests that the p-value selected features (average unweighted clustering coefficient, characteristic path length, transitivity, right dorsal amygdala node degree, right dorsal amygdala node strength, and right dorsal amygdala closeness centrality) were not independent of one another. This is unsurprising, especially considering that three of the features correspond to the same ROI.

Conclusion

This paper identifies a potentially significant finding of reduced transitivity in subjects with mild traumatic brain injury, suggesting that cranial insult may lead region-based brain cliques to break down. However, there is no evidence to suggest that the graph theory metrics investigated in this paper serve as strong predictors of traumatic brain injury on their own.

Limitations:

The study size is relatively small (n=104), giving the current analysis low statistical power. As such, there may be features which truly differ between mild TBI and healthy control groups that were not identified as significant. In the same vein, it proved impossible to learn a complex relationship between rsfMRI brain image graph metrics and brain injury with such a small dataset.

Furthermore, numerous graph theory metrics which might prove to be useful descriptors for injured versus healthy brains cannot be calculated using the current standard processing methodologies. For example, weighted shortest path length between nodes could theoretically be calculated, but with edges calculated according to the standard methodology (edge weights = Fisher's z values), a shorter weighted path length between nodes would simply indicate that those nodes were connected via an uncorrelated path, which is not intuitive.

Future Directions:

While the current machine learning results are not particularly promising for the goal of differentiating TBI patients from HC patients, further steps could be undertaken to try to achieve an accuracy better than the naive case. Firstly, principal components analysis (PCA) could be applied to the data as a preprocessing step to see whether arbitrary axes of maximal orthogonal variance yield better predictions. PCA would allow models to learn fewer parameters, which is essential given the small dataset size, but at the same time retain variance from all or most of the original features.

An undersampling cross validation class could be implemented to reduce overfitting to noise during cross validation. Because validation data is oversampled, some bias is introduced to the validation accuracies, where oversampled samples are treated as more important to the accuracy score.

A separate machine learning problem could be investigated with Brainsport's CBF rsfMRI data from time point 2 (T2) and time point 3 (T3). A multiclass classifier could be trained to differentiate the three time points from one another. Additionally, time series features could be generated, and used to contrast the two later time points.

Lastly, boosting methods were not employed in the current analysis, due to a focus on solving data deficiencies contributing to other model's lack of success. AdaBoost and XGBoost should be considered as further models to train and tune.

Works Cited

1. URL: (https://www.frontiersin.org/articles/10.3389/fnins.2018.01018/full)
2. URL: (https://www.liebertpub.com/doi/10.1089/brain.2012.0073)
3. URL: (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4961028/)
4. URL: (https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0178798)

Code for this project can be found at: https://github.com/danielfrees/TBI_BrainGraph_ML

Note: Most of the important work for this project occurs in T1_Explore.ipynb, and I added HTML markups and a table of contents to guide navigation through the code and results in this file.

Scripts are currently being organized in the scripts/ directory under this repository, for the further analysis I plan to conduct at the start of next quarter.