Predicting of Handedness with Resting-State fMRI Data

Vicky Li yil164@ucsd.edu
Zhuoxuan Ju zju@ucsd.edu
Anastasiya Markova anmarkova@ucsd.edu
Armin Schwartzman armins@ucsd.edu
Gabriel Riegner gariegner@ucsd.edu

Abstract

Handedness is a key aspect of human lateralization and has been linked to distinct neural connectivity patterns in the brain. In this study, we aimed to classify handedness using functional connectivity features derived from the Human Connectome Project (HCP) Functional Magnetic Resonance Imaging (fMRI) data with the help of machine learning tools. We used Pearson's correlations to investigate the connectivity through correlations between fMRI signals and as one of our inputs for our models. We implemented multiple machine learning models, including Support Vector Machines (SVM), Logistic Regression, Polynomial Regression, Stochastic Gradient Descent Regressor and Convolutional Neural Networks (CNNs) to analyze and predict handedness based on brain connectivity. Given the strong class imbalance and small datasets, with right-handed participants significantly outnumbering left-handed ones, we experimented undersampling and oversampling, and we are currently working on stablizing the model performances to a specific range.

Code: https://github.com/vickyli1015/Public-DSC180B-Functional-Connectivity-Project

1 Introduction

Lateralization is the idea that some human functions and behaviors are specialized to one area of the brain. Perhaps one of the most mundane examples of this is handedness. Handedness—whether a person is left- or right-handed—is an essential characteristic linked to various cognitive and neurological traits. Right-handed people comprise about 90% of the world population, while left-handed people are only about 10% of the world population. Majority of research is focused on right-handed people as well, hence generalized conventions about brain functions could be different for left-handed people as their brain would be lateralized differently. Understanding how neural connections and activations are correlated with handedness is still an area of research though.

One way to investigate these differences is through resting-state functional Magnetic Resonance Imaging (rfMRI), which captures spontaneous brain activity and connectivity patterns. However, the relationship between resting-state functional connectivity and handedness remains an open research question. With machine learning techniques, there is an opportunity to classify individuals as left- or right-handed based purely on their rfMRI-based neural signatures. While previous studies have explored this classification, there is room for improvement in terms of generalizability and feature selections.

With machine learning techniques, we had the goal to classify subjects as left-handed or right-handed based solely on their rfMRI signals. Our main goal was to recreate and improve upon already existing results in handedness classification from resting state fMRI data. Along the way we hoped to uncover potential networks in the brain associated with handedness.

1.1 Literature Review

Previous studies in rfMRI were able to identify motor systems, specifically primary motor cortex (M1), supplementary motor system (SMA) and dorsolateral premotor cortex (PMd), as key features in handedness classification (Pool et al. (2015)). These researchers were able to reach an accuracy of around 86%. Using these regions as seed regions to derive correlation with the rest of regions. This study had only 18 right and 18 left hand participants, so although they were successful it is unclear how generalizable this would be to a larger dataset. They tested that motor regions are the ones that are predicting handedness by testing visual and auditory regions as well. They showed that inferior frontal gyrus as seed region does not have significant differences between left and right handed people. Similarly, visual cortex correlations did not show significant differences between left and right handed people. Their high accuracy could be linked with the fact that it was shown that right handed people had a stronger interhemispheric link between M1 and PMd compared to left handed people.

While Pool et al. (2015) focused on moter-related regions in classifying handedness, other studies have examined the broader functional network properties that differentiate left and right handed individuals. Li et al. (2015) extended this investigation by analyzing small-world properties in resting-state functional brain networks, highlighting that handedness is associated with differences in network efficiency. They found significant alterations in nodal efficiency in left-handed individuals, particularly in regions beyond the moter system, such as the anterior and median cingulate gyrus, middle temporal gyrus, angular gyrus, and amygdala. These findings suggest that handedness-related brain differences are not limited to motor regions but also involve areas related to cognitive and emotional processing.

As research has expanded beyond motor regions to explore network-level differences in handedness, new computational approaches have emerged to better capture the complexity of brain connectivity patterns. Meszlényi, Buza and Vidnyánszky (2017) explored deep learning methods for decoding individual brain states from resting-state fMRI (rs-fMRI), demonstrating that CNNs outperform traditional machine learning models by capturing complex spatial patterns in functional connectivity matrices. While their study focused on general brain state classification, it highlights the potential of deep learning for brain-based classification tasks. However, their work did not address handedness prediction specifically since their goal was to predict Autism, nor did it consider class imbalance issues, which are crucial in our dataset where left-handed individuals are underrepresented.

2 Data

For the project this quarter, we used the resting-state fMRI data from the Human Connectome Project (HCP). Since our main focus was on connectivity patterns between brain regions over time, we selected the HCP_PTN1200 dataset, which captured brain activity and regional connectivity represented across time.

The datasets included brain activity data across various numbers of regions, ranging from 15 to 300 brain regions. The activity of each region was represented as a continuous time series, and more regions provided finer details. For our analysis, we decided to use and focus on 50, and 100 brain regions for 1,003 subjects to compare and contrast the results of connectivity at varying region levels. Each subject's data consisted of a single file containing 4,800 time points, representing one hour of resting-state brain activity. This hour of data was divided into four 15-minute rfMRI sessions, equivalent to 1,200 time points per 15-minute scan. The data we used measured brain activity across 15, 50, and 100 regions determined via ICA. When we used 15 regions, there were no centroids located in the front lobe, while when 100 regions were used, there was a significant overlap in the posterior region of the brain. Note that this does not mean that frontal lobe data is not recorded, but rather that it is a part of a larger region whose centroid is located somewhere else farther away from the frontal lobe.

The ICA parcellated regions combine brain activity from both hemispheres, which is incompatible with predicting handedness. Thus, we selected 172 subjects' raw data files and extracted 360 cortical regions, 180 per hemisphere, using the Multi-Modal Parcellation (Glasser et al.) Furthermore, we extracted 712 subcortical regions, 356 per hemisphere, using the Cole-Anticevic Brain-Wide Network Partition. For each of the brain regions, we have 3600 time points. All of these data were extracted and preprocessed by our TA Gabriel

Riegner.

From the same study, we used restricted data on handedness. Handedness is assessed using the Edinburgh Handedness Index, which ranges from -100 to 100. Negative values correspond to left-handedness, while positive values correspond to right handedness. The index is derived by answering a questionnaire, which asks participants which hand they use to perform different tasks such as:

- Doing Fine Motor Tasks: writing, drawing, throwing, threading a needle, etc.
- Using Eating & Kitchen Tools: Knife (without fork), Spoon, Knife with fork

When we split data into left- and right-handed people, we discovered that our data consists of 13 left-handed people and 159 right-handed people.

3 Methods

Since our dataset consists of 13 left-handed people and 159 right-handed people, we chose to try to oversample left-handed people in order to balance the weights of the two groups. We used Peason's correlation transformed through Fisher's z-scores to derive correlations between voxels of different regions. This standardization allows us to have comparable values between different correlations, so that correlation coefficients are standardized. We also derived partial-correlations, which show more direct correlations between regions since they calculate linear correlation between regions while also taking into account the influence of other regions in the network. We include these correlations as features in our models. While we tested all areas of the brain, we primarily focused on the motor and sensory areas such as Primary Motor Area (M1) and Dorsolateral Premotor Cortex (PMd). Prior studies also utilized gender as a covariate, which we include as a feature in our model as well.

We focused on 3 main models for our prediction: Classification, Regression, and Neural Networks.

3.1 Classification

Since our prediction variable, EHI, is continuous we first make it discrete by thresholding from 0 to 25 to differentiate between left and right handed people. In the paper by Eva-Maria Pool et al., 2015, the authors were able to achieve 86% accuracy with the SVM model. We utilized Support Vector Machines, in order to predict the label of the left or right hand. This is what people currently use the most for fmri data, and we want to utilize it first as it is common among academic papers. After using the SVM, we use a logistic regression model which allows us to experiment with the prediction outcome more freely by tuning the threshold of the sigmoid function which means that logistic regression allows us to be more flexible in tuning the threshold and hyperparameters.

3.2 Regression

In regression we primarily focused on polynomial regression which allows us to fit non-linear data when we fit the regression to a specific degree, which we can tune to our training data. Another model we considered is stochastic gradient descent regression allows us to iteratively minimize the L2 loss, similar to the SVM. These models rely on the fact that our data is somewhat linear, which means that they won't achieve a high performance unless this assumption is true. We have not seen many applications of linear models on fMRI data, but we still wanted to try them since they are simplest continuous models and would work well with our prediction coefficient.

3.3 Neural Network

In addition to regular classification and regression, we also wanted to explore the Convolutional Neural Network (CNN) that Meszlényi discussed. One of the key challenges was the severe class imbalance, with right-handed subjects significantly outnumbering left-handed ones. To address this, we experimented with SMOTE-based oversampling on the training set, as well as class-weight adjustments during CNN model training. However, with or without applying SMOTE, we encountered high performance fluctuations across different training runs. Despite applying SMOTE-based oversampling and class-weight adjustments, test accuracy varied significantly between runs. The balanced accuracy was around 0.5 in some cases, showing that the CNN may not be properly differentiating between left- and right-handed individuals. We plan on doing further analysis and explore more hyperparameter tuning to determine whether the model is truly learning functional connectivity differences or is overly sensitive to noise in our data.

4 Appendix (Copy of Project Proposal for Reference)

Lateralization is the idea that some human functions and behaviors are specialized to one area of the brain. Perhaps one of the most mundane examples of this is handedness. Handedness—whether a person is left- or right-handed—is an essential characteristic linked to various cognitive and neurological traits. Right-handed people comprise about 90% of the world population, while left-handed people are only about 10% of the world population. Understanding how neural connections and activations are correlated with handedness is still an area of research though. Resting-state functional magnetic resonance imaging (rfMRI), which measures spontaneous brain activity while the subject is not engaged in specific tasks, has emerged as a powerful tool for exploring brain networks. By leveraging machine learning techniques, we propose to classify subjects as left-handed or right-handed based solely on their rfMRI signals.

We propose to leverage our existing dataset of resting-state fMRI data from 1,003 subjects, spanning 4,800 timepoints, to predict whether the individual is left or right handed. By leveraging newly available behavioral and demographic data, we explore how patterns in functional connectivity across the brain can serve as predictors of handedness. By exploring and predicting how functional connectivity and neural activity relates to handedness, we can enhance our understanding of lateralization in the brain, shed light on conditions like cognitive skills such as language, and psychiatric disorders such as attention deficit hyperactivity disorder (ADHD) and depression. Non-right-handedness has also been suggested to be linked to early life factors and various neurodevelopmental and psychiatric disorders, such as autism and schizophrenia. These are mostly from correlational studies rather than based on neurobiological analysis, which is what we aim to do. Furthermore, our approach could inform biomarker development for clinical applications.

4.1 Narrow, careful problem statement

In our quarter one project we explored methods that can help describe brain connectivity. We performed various statistical analysis on correlation matrices across subjects, time and brain regions. We performed PCA to find networks of brain regions that are connected and perform specific brain functions. We are sliding time windows to see how connectivity changes over time, and across distances in the brain. Through our analysis we were able to observe a variety of interesting patterns between subjects. We will apply our knowledge of existing patterns to extract features from the fMRI data and use them for our prediction of white matter. One useful pattern we have already identified is that using less brain regions gives us more insights into correlations within hemispheres. Some of the more basic features could be features that we already extracted such as squared average correlation and absolute average correlation. In quarter one project we found it difficult to compare the subjects in a quantitative way, although it was evident that connectivity patterns vary greatly between subjects. Some subjects tend to have higher correlation scores than others. Some tend to have interesting patterns in their data such as two concentrated clusters of correlation, or correlation decreasing as the brain regions are further apart. However, these are not generalizable but rather singular observations. We want to apply the exploratory findings from Quarter 1 to create a more tangible conclusion about the way the brain functions in Quarter 2. We will explore three main directions. One direction is to extract general statistics that describe our connectivity correlation matrices and use those as features to predict left and right-handedness. Second direction is to utilize the results from PCA, and use the connectivity relationships from PCA and utilize those as input vectors into a neural network and other classification models like Support Vector Machines that can make predictions. Additionally, we can utilize connectome based predictive modeling, to use our correlation matrices in order to predict a certain feature. Prior research has already shown that this sort of classification is possible. In Chormai et al. (2022), the researcher conducted the study that used resting state fMRI data to predict left and right handedness. Overall, they were able to reach a good AUROC score of up to 0.7243. We believe that their results are reproducible, since they concluded that the most predictive features were related to functional connectivity. We believe that we can build on top of their results and try additional models in order to see if we can reach higher accuracy. Although the challenge will always be an imbalance in data, since only 10

4.2 Primary Output

The primary output of this specific project is the estimation and prediction of the handedness of the subject from their resting state fMRI data. This project will utilize the resting state fMRI data for each person to create a model that can accurately predict whether the subject is left-handed or right-handed only providing their resting state fMRI.

At first, we will write a reproducible paper that details how to explore and extract features from fMRI data and train the models that can predict subjects' handedness. We will discuss our feature extraction, model selection, and fine-tuning the model process in the methods section of our paper. We will also discuss our model performance and its strengths and weaknesses in the results section of the paper. We will include any interesting generalizations and conclusions that our model produces in the discussion section. By the end of our process, we hope that we are able to talk more about characteristics of a brain which can potentially be explained by the connectivity generated from the resting state fMRI data .

Additionally, we are going to create a website that contains our findings and results of this project. This website will demonstrate all of the information from our paper, as well as act as a more interactive way to view our results and how the model works. We will visualize the brain connectivity image again and form a comparison and highlight of the key connectivities or regions that help us to determine the handedness of the subject in that specific testing subject's brain.

4.3 Justification

With current studies about the differences between left-handed and right-handed people's brain and connectivities, this project provides a model that analyzes and summarizes the pattern and pairs of connections using the resting state fMRI data which did not previously been considered as the most direct way to measure. Specifically, we all know that the researchers prefer to use a task involved data of the brain to study handedness, and we noticed that resting state fMRI data also enter the conversation, but are not being effectively and thoroughly utilized in this particular field. So our topic already had numbers of researchers contributing and focusing to it, and we decided to go deeper into that field.

As for the data, we are using the data from the Human Connectome Project (HCP) which is a database for fMRI, and this database was worked on by a lot of researchers. For our project, we will use the resting state fMRI data and the subject's handedness and demographic data for our project. The resting state data contains the brain signal for each subject within the study and records both the group ICA and time series of the subject's data which are divided by different scales of brain regions.

From all the descriptions above, we believe our project will be successful. Specifically, we can get hands on a lot of related projects and research, and we already learned some essential knowledge from our previous project. As to the topic, there have been a number of relative projects that demonstrate some correlations and relationships about the handedness towards the brain activity, and our project pushes the concept further and explores different possible alternative explanations instead of staying in abstract descriptions. This decision offers us the backbone toward the success of our project. For the data aspect, we believe that the data's quality, scholarly professionalism, and the broad variability in different scales of regions provide a reliable and high quality characteristic of our data, which emphasize the success of our project.

5 Contributions

Anastasiya Markova:

- Week 1: Read research papers related to the topic, which were used as the basis for our research.
- Week 2: Worked on the regression models including Stochastic Gradient Descent, and Regression Models
- Week 3: Obtained DSMLP access, moved all of the data to DSMLP and worked on balancing the dataset with sklearn.imbalanced
- Week 4: Worked on feature extraction such as correlation matrices, looking for specific regions of interest, partial correlations and standardizing through Fisher's z-score.
- Week 5: Found incompatibility between our dataset and out objective. Obtained new dataset, conducted basic exploratory data analysis on it. Began to process the data set to fit with our existing code for models.

Vicky Li:

- Week 1 (Literature Review): Read papers about CNN and motor areas relevant to handedness
- Week 2 (EDA) :
  - Plotted the distributions of left- and right-handed subjects, found significantly skewed distribution and sizes.
  - Conducted hypothesis test on the medians, found significantly different medians too.
  - Applied different thresholds (+/-25% and +/-40%) to handedness and plotted the distributions, which were both found out to be skewed.
  - Plotted the distribution of basic demographic features across handedness classes.
- Week 3 (EDA/Preprocessing): Generated visualizations per demographic feature of interest, showing their distribution and relation to handedness, Created correlation matrices between the brain regions of interest (motor) and the handedness of the subject for both d50 and d100 Tried transforming Handedness distribution to prepare for modeling
-

Zhuoxuan Ju:

- Worked on the calssification models including primarily Support Vector Machines.
- Balancing the dataset through undersampling and oversampling
- Read research papers related to the topic.
- Looked for regions of interest using feature selection algorithms

References

Chormai, Piyapong, Yang Pu, Heng Hu, Simon E Fisher, Clyde Francks, and Xiang-Zhen Kong. 2022. "Machine learning of large-scale multimodal brain imaging data reveals neural correlates of hand preference." *NeuroImage* 262, p. 119534. [Link]

Li, Meiling, Junping Wang, Feng Liu, Heng Chen, Fengmei Lu, Guorong Wu, Chunshui Yu, and Huafu Chen. 2015. "Handedness- and brain size-related efficiency differences in small-world brain networks: A resting-state functional magnetic resonance imaging study." *Brain Connectivity* 5 (4): 259–265. [Link]

Meszlényi, Regina J., Krisztian Buza, and Zoltán Vidnyánszky. 2017. "Resting State fMRI Functional Connectivity-Based Classification Using a Convolutional Neural Network Architecture." *Frontiers in Neuroinformatics* 11, p. 61. [Link]

Pool, Eva-Maria, Anne K. Rehme, Simon B. Eickhoff, Gereon R. Fink, and Christian Grefkes. 2015. "Functional resting-state connectivity of the human motor network: Differences between right- and left-handers." *NeuroImage* 109: 298–306. [Link]