BIOSTAT-625 Group 1 Project Proposal

Predicting Dementia from MRI Imaging

Michael Miller, Sabir Meah, Liyang Yuan

Dataset: We decided to use an Alzheimer's MRI imaging dataset from Kaggle. The data includes a total of 6400 MRI images from individuals classified into four stages: no dementia, very mild dementia, mild dementia, and moderate dementia. Each MRI result is a 128x128 .jpg image. The images have been preprocessed so that they are of a standard form and orientation, but we still will need to do the step of processing the data and gray scale standardization from .jpg images into a tabular format suitable for standard R functions.

Methodology: We intend to use convolutional neural networks (CNNs) to predict the individual's dementia status. CNNs and their extensions are the state-of-the-art methods for classifying images that contain translation-invariant features. Convolutional layers use far fewer parameters than fully connected layers and do not require data augmentation to feed spatial features in as inputs, which greatly reduces overfitting.

One challenge will be to address the ordinal nature of the classes. If we treat the classes as unordered categories, we will sacrifice some predictive power because we will not be using our prior knowledge of the ordering of the class. In other words, the predicted probabilities from the softmax output should be monotonically increasing or decreasing in the order of the categories. One potential method is to predict an underlying continuous variable and then learn the cutoff values that correspond to each of the classes. Existing methods do not appear to account for this structure, so our goal is to construct a novel CNN to predict this type of data particularly well. We will use the Keras R package.

Dataset citation: Kumar, S., & Shastri, S. (2022, March). Alzheimer MRI Preprocessed Dataset. *Kaggle*. Retrieved November 20, 2022, from Kaggle. https://www.kaggle.com/datasets/sachinkumar413/alzheimer-mri-dataset