Diagnosing Major Depression from fMRI Data

Christian Howard Sameer Manchanda Ryley Higa

Abstract

A set of classifiers were built that could be used to predict if an individual has Major Depression given fMRI data about them performing a set of tasks. To accomplish this task, a dense fMRI dataset was coarsened via a local voxel averaging scheme and then features were discovered using Principal Component Analysis (PCA) and Non-negative Matrix Factorization (NMF). Using the new feature representations, classifiers were constructed using the following approaches: Random Forests, Linear Discriminant Analysis (LDA), and a Deep Neural Network (DNN). All of the above classification methods had high testing accuracy. This high accuracy was interesting and appears to stem from the feature representation causing the reduced data to be very separable with respect to their classes.

What is fMRI data?

fMRI stands for Functional Magnetic Resonance Imaging. The idea of how this works is that tissue is placed within a strong magnetic field. To help capture a picture of the tissue, radio waves are sent into the tissue. Blood that is deoxygenated will then reflect the waves in a manner that can be processed by sensors and algorithms to generate a fMRI dataset. Figure 1 shows example visualizations of some subset of the fMRI data used within this problem.

Dataset Details

The dataset used within this analysis was obtained from OpenfMRI. It contains data from 19 individuals with Major Depressive Disorder and 20 individuals with no mental illness. fMRI data was then collected for all these individuals while they did a set of tasks related to listening to different clips of audio. The resulting dataset is a time series information of Blood Oxygenation at a dense point cloud of locations within the patients' brains.

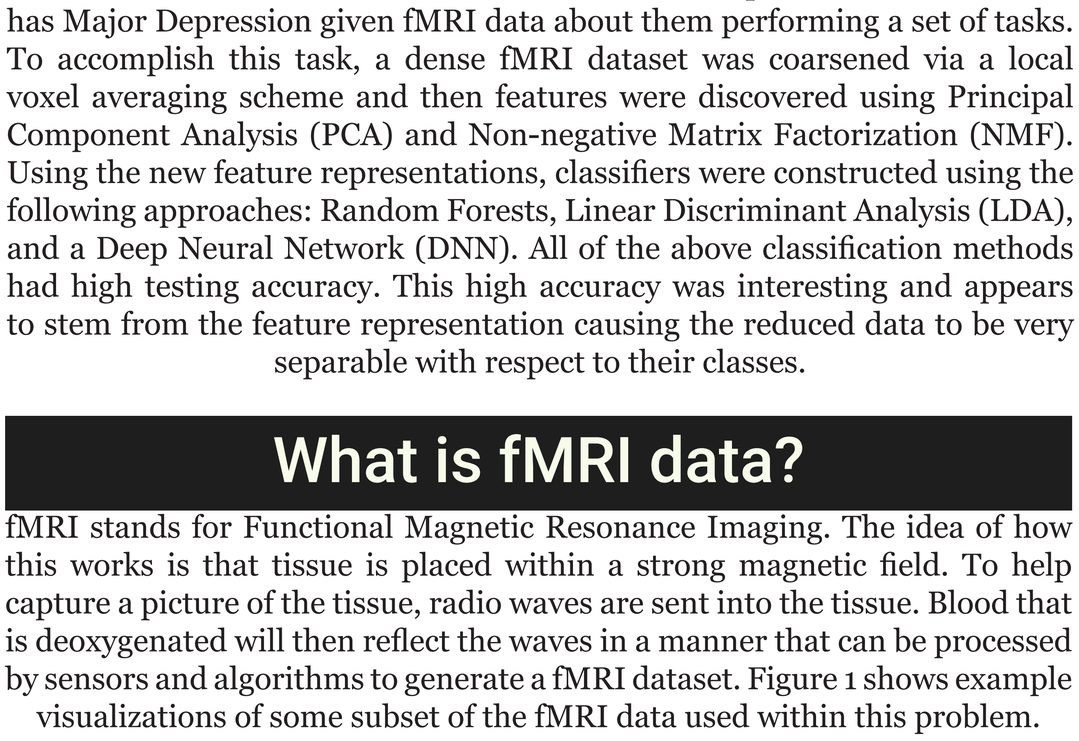


Figure 2. Flowchart of data processing and classifier building

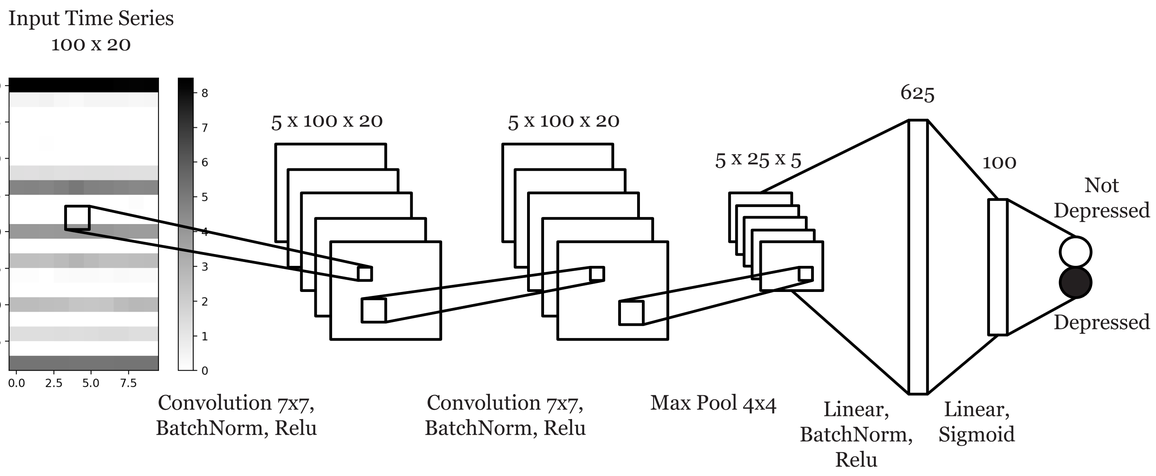


Figure 3. Deep Neural Network Configuration

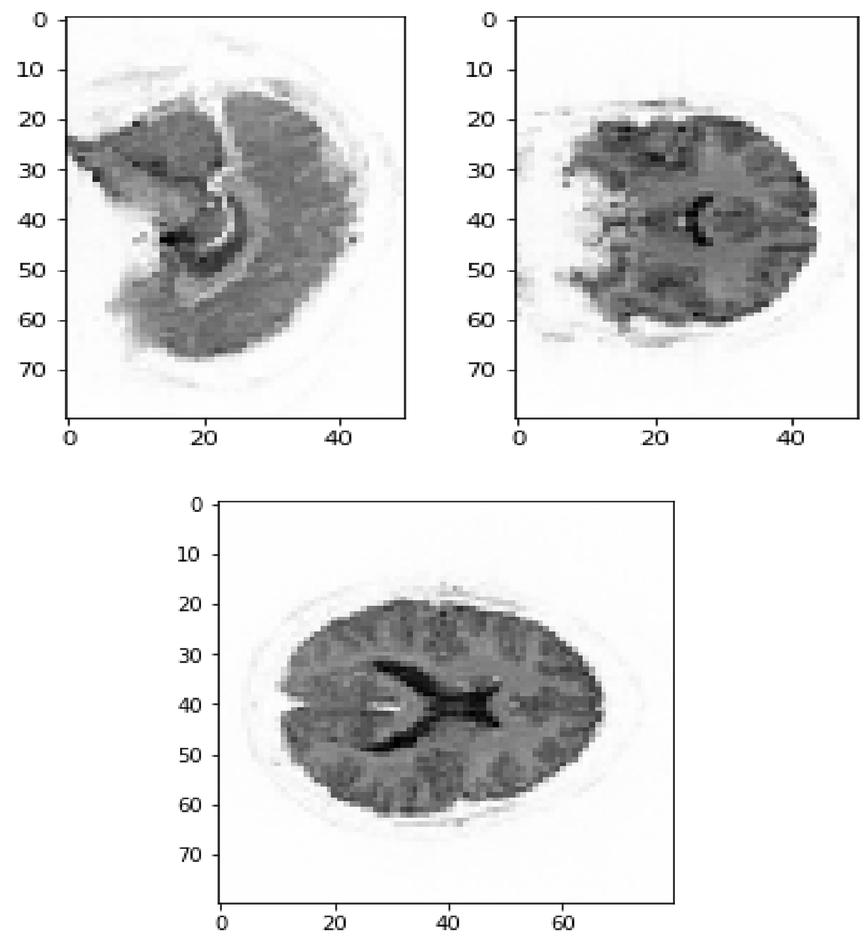
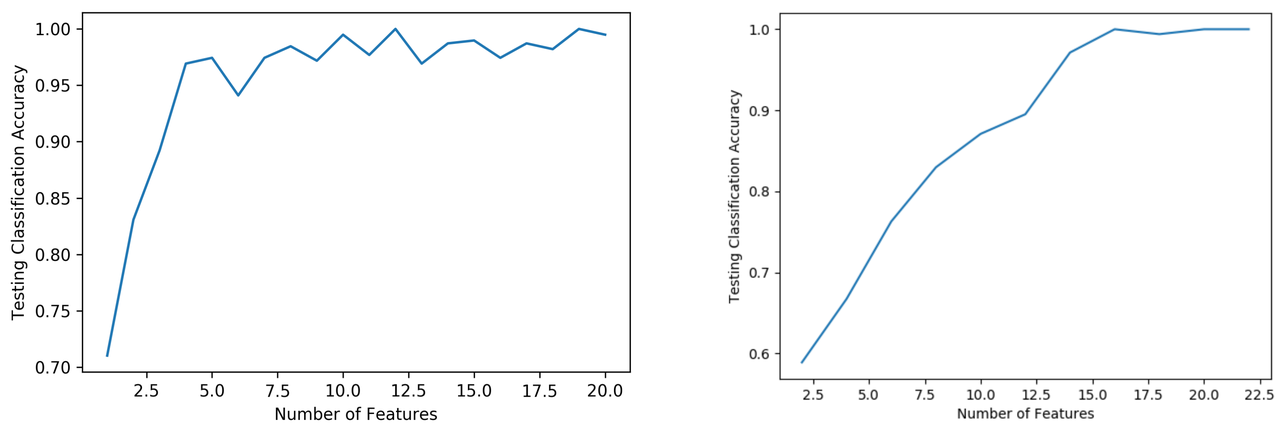


Figure 1. Image slices of the Intensities of Blood Oxygenation levels of some brain at the initial time

Results

After performing the necessary coarsening and dimensionality reduction, a set of models were built for predicting if someone has Major Depressive Disorder based on their fMRI data. All models performed well but, as Figure 4 shows, their success for a given number of features was dependent on the modeling approach. Random Forests proved to not only perform well for the provided NMF features, but was very fast to compute. LDA performed well using PCA feature as the number of PCA features increased and was also very efficient to compute. Now while the Deep Neural Network required more computational complexity to train, it showed great performance as well. As can be seen in Figure 4, LDA required the most number of features to achieve greater than 90% classification performance while the Random Forests and Deep Neural Network needed only 5 or more features to achieve high accuracy.



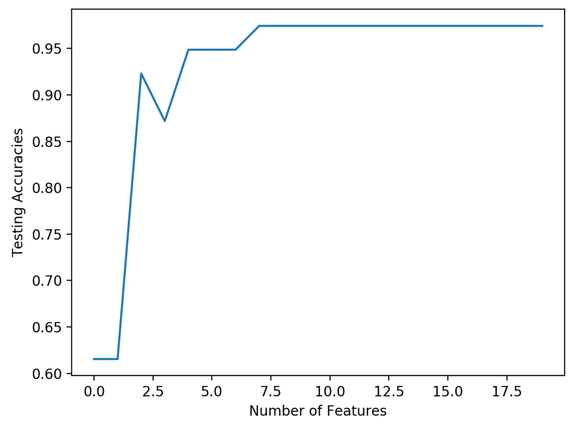


Figure 4. Testing Accuracy as function of number of features. a) Random Forests with NMF Features. b) LDA with PCA Features. c) Deep Neural Network with NMF Features