Accelerating the Diagnosis of Depression with Computer Vision

Shubhom Bhattacharya
sb2287@cornell.edu
Department of Computer Science
Cornell University

***Abstract*—Mental health awareness is a growing movement, but prevention and diagnoses of mental health disease is difficult. Unlike many other parts of the body, the human brain is not understood to the extent that direct relationships between external stimuli and temporary moods, long-term depression, anxiety, etc. are fully understood. Deep learning algorithms provide a way to extract high-level information from images, and thus present a promising method to aid in diagnosis of mental health disorders. In this project, we applied and evaluated several neural network architectures to the binary classification problem of depression diagnosis in the fMRI dataset openneuro-ds000171 and present several promising results.**

I. Introduction

Mental health awareness has increased in recently history as suicide rates continue to grow across demographics [1]. Today, suicide is the second leading cause of death in individuals between age 10 and 34 [2]. Suicide often affects not only the individual, but also the individual's immediate family and friends, community and society as a whole. Current research in depression and mental health aims to mitigate the risk of suicide by diagnosing mental disorders before they exacerbate and to treat individuals with mental disorders using techniques such as hormonal therapy and meditation [3]. As public attitude shifts towards viewing depression as a mental condition, finding physiological effects of the disease becomes useful. In particular, understanding the depressed brain allows doctors to improve treatment by searching for regions-of-interest with high brain activity and monitoring the effectiveness of administered drugs. In law, the potential physiological definitions of behavioral diseases eliminates "grey areas" that otherwise allow insurance companies to abuse these ambiguities by denying coverage for mental illnesses. In science, mental health issues provide a novel perspective into the workings of the human brain in ways that may not have been known before – are depression and normal feelings of sadness physiologically distinguishable? This project aims to investigate the use of computer vision techniques to study fMRI scans in 39 patients divided into two classes: 19 with a history of chronic depression and 20 with no known history of chronic depression. Each patient, for 5 trials, is subject to positive and negative musical and nonmusical stimuli and is thoroughly scanned. Using this data, we take a machine learning approach to diagnose depression in a given fMRI scan. For background, Functional Magnetic Resonance Imaging (fMRI) is a technique in which hemodynamic response to the brain is used to map brain activity. More specifically, blood releases oxygen causing neurons to fire at a higher rate. On one hand, oxygenated blood contains oxyhemoglobin which is a diamagnetic. On the other hand, deoxygenated blood contains deoxyhemoglobin which is paramagnetic. The relative difference in these concentrations can be traced by magnetic properties in a process called Blood-Oxygen-Level-Dependent (BOLD) contrast, lending insight into brain activity[4].

II. Approach

A. Prior Research

Prior research in this area has used Support Vector Machines to process fMRI imaging data [5]. This is based on the belief that to fMRI data is inter-subject and consists of many variables, so a multivariable and brain response model-free method is required to make diagnoses [9]. We relax this assumption and propose a new method using convolutional neural networks as the first step to processing fMRI imaging data. While these are similar in difficulty in terms of implementation, there are additional challenges with CNNs including a need for abundant training data, large computational resources, and , as well as the lack of prior research and relatively small dataset. There is also the added difficulty of eventually preprocessing managing large amounts of patient data, which we resolve by building a fMRI processing pipeline and creating a 3-dimensional neural network that is both hyper optimized and light-weight in order to provide timely output.

B. Dataset

ds000171 is dissimilar from [5], which uses diffusive tensor imaging as opposed to fMRI. Rather than manually scanning pertinent volumes and performing analysis of chosen regions of interest, we propose using a CNN-based algorithm to automatically find and extract useful features. As mentioned previously, 39 patients (19 depressed and 20 non-depressed) are subject to musical and nonmusical stimuli over 5 trials while their brains are scanned. It is important to note that the depressed patients are experiencing a depressive episode at the time of scanning.

C. Objectives

Like previous studies, our general objective is to use computer vision techniques to determine patterns of brain activity that indicate depression. Our approach differs from

predecessors by the algorithms applied. Our first goal is to determine if there is a CNN architecture that can show some improvement over a baseline metric as well as other models used in previous literature, such as the SVM. Furthermore, we hope to compare newer model architectures to assess whether these neural networks have added inference ability.

III. Preprocessing

Traditionally, fMRI scans have four associated dimensions (a 3-dimensional brain volume over a period of timesteps lasting the duration of the scan). This time series data is preformatted in a Brain Imaging Data Structure (BIDS) format. We use the BIDS format to maintain compatibility with other models for late stage testing. We primarily use the fmri prep package on brainlife.io to preprocess our data. Brainlife.io makes available data and analyses software that allows us to run various algorithms on our datasets through built pipelines. We use fmriprep on our raw data to output standard preprocessed BOLD data. This includes motion correction, skull stripping, brain region segmentation and reorientation. We also obtain a derivatives csv file which indicates the estimated amount of motion at each frame, to be used in analyses as a nuisance covariate for later data processing. Next, we take each preprocessed brain volume and examine three different angles at various time intervals. For each time interval, we save the middle slice at three different angles that will eventually be fed into our convolutional neural network. This greatly decreases the possible regions of interest and narrows the learning problem to the available image regions instead, which represent the largest possible view of the brain in 2-dimensions. We justify this by making the assumption that blood oxygenation rate does not change between the depressed and control subjects, and that the oxygenated regions are located closer to these volume axes because the outer regions of the brain are likelier to be grey matter or peripheral organs. To allow some flexibility in the data, we add Gaussian noise with variance of 5 pixels to the precise location of the input data, so some regions of interest that are near the volume axes are represented in the data.

Figure 1 shows examples of three mid-axial scans at t=0, t=52, and t=104 (row 1, 2, and 3 respectively). We can see that the oxygenation does not vary to the extent that pooling operations in a CNN would likely act as a high-pass filter to obfuscate the details varying between the timesteps.

IV. Results

A. 3-D CNN

3-D CNNs presented a logical way to retain correlations of brain activity by voxel without sacrificing some degree of spatial information. After preprocessing, we first tried to implement a 3-dimensional convolutional neural network to see if we achieved low training error, but found when training the model that it required too much time and memory, even with a relatively small number of 3-dimensional filters (less than 64 in any layer). Due to the added dimension the number of filters grows very quickly given the relatively large volume size (padded to 70 x 70 x 70). Even when taking a single average of the entire time-series volume, computational resources were not present to perform meaningful training, so this first foray was abandoned.

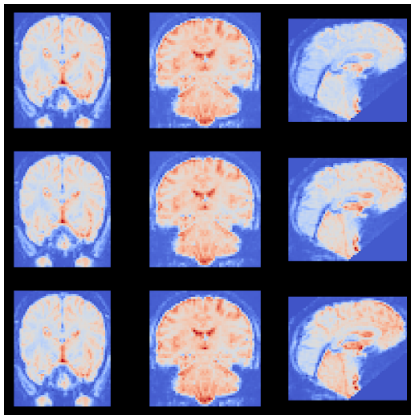

Fig. 1. Figure 1: Mid-axial volume scans

B. Initial 2-D CNN

As a next attempt at the binary classification task of depression diagnosis, we use a model architecture similar to AlexNet with 7 convolutional layers and 3 fully connected layers trained on the mid-sagittal, mid-traversal, and mid-coronal scans (essentially stacked into an image with 3 channels) with added Gaussian noise off the axes to increase the training data variance. We were able to achieve a decrease in empirical training loss, indicating visible optimization results. However, the model accuracy was exactly the baseline (51%) so significant results were not achieved, though a subset of the training data was ready for use and more complex models were not considered. Interestingly, at first, in lowering the batch size and learning rate, we were able to see much greater training and validation accuracy results (90%+ and 80%+ respectively) but further analysis revealed that the network had learned to match the brain shapes between runs in the training set and validation set. To ensure this did not happen, we converted the training and validation set to be patient-based; i.e. 30% of patients were chosen to be in the validation set instead of 30% of all scans present. This enabled models to learn high-dimensional fMRI patterns as opposed to simply overfitting to patients' brain structure. One main merit in our neural net was how fast we were able to train. On a parallelized-TPU framework, though several trials, training took on average 150 epochs and 1000 seconds. This is much quicker than our 3-dimensional CNN, so we intend to explore this idea further to achieve lower training error. However, we found this naive architecture was not sufficient for improving over guessing at random after the repartitioning of the training and validation data. We hypothesized the first attempt's failure on the lack of a sufficiently rich feature space. We also believed that filtering and pooling with the chosen kernel sizes was risky, as the scan itself is noisy to begin with and we were not certain of what the model should be looking for a priori. As a result,

we applied several candidate architectures that have shown remarkable success in image classification tasks to address our original architecture's purported deficiencies.

C. Deeper Networks

The first of these candidate architectures, VGG-19 [10], simply added more layers and batch normalization. The added number of parameters we hoped would be better able to adequately learn the rich feature space, and batch normalization, as commonly used in deep learning, was expected to reduce artifacts of increased variance from the network weights being multiplied. This architecture showed modest improvement over the original CNN. This achieved training accuracy around 65% and validation accuracy 58%.

Inception Modules The next of these architectures incorporated inception modules[11]. Here, we wanted to see whether the ambiguity in filter sizing and pooling could be addressed. This proved to be the most effective of the methods attempted, reaching training accuracy of over 85% and validation accuracy near 70% using GoogLeNet. Interestingly, the Inception-v3 architecture [12]was actually worse at the classification task, achieving only the baseline (51%) training accuracy. This however may be an artifact of poor hyperparameter tuning.

D. Residual Connections

The last of these architectures utilized residual connections. Using a very deep network, Resnet-101, we wanted to assess whether residual learning could improve on the non-residual and relatively shallow networks used previously. After over 100 epochs of training, training accuracy did not reach over 57% and validation accuracy hovered near the baseline.

V. Conclusions and Next Steps

The next steps in this project are to increase model complexity and augment the training data in order to achieve lower training error. Once we achieve lower training error we can statistically compare our approach against the traditional SVM approach and then use more complex models. We find that the GoogLeNet architecture shows promise and significantly improves from the baseline accuracy. With more time and resources to perform extensive hyperparameter tuning, we believe that, on this dataset, the algorithm could demonstrate some knowledge of physiological responses to depression. We claim this by assuring validation subjects are separated from training subjects and our analysis involved all stimuli. However, before concluding the model learned depression diagnosis from fMRI, it is important to note that physicians' opinions should be taken into consideration as there may be tangentially related phenomena that may affect our results we are unaware of. One corollary to this project is to incorporate the region-of-interest correlation matrix from BOLD preprocessing into the learning algorithm to see if our learning algorithm can be guided by fMRI correlation matrices. Another interesting project would be to incorporate other biological metrics into our analysis. In modern day diagnoses, metrics such as pulse rate to varying stimuli, the rate of secretion of hormones and various psychological stimuli are readily available for patients seeking depression diagnoses, so we hope to incorporate these data sources into our analysis. Overall we are satisfied with the results of this project in finding some physiological depression diagnostics using fMRI. Furthermore, we were able to do this using a novel, efficient preprocessing method aimed at maximizing exposure to possible regions-of-interest while effectively reducing the dimensionality of the fMRI scan by 2.

Please contact the author for access to data or code.

References

[1] Suicide Statistics. (2019, April 16). Retrieved from https://afsp.org/about-suicide/suicide-statistics/.

[2] Depression Statistics. (n.d.). Retrieved from https://www.dbsalliance.org/education/depression/statistics/.

[3] Bilodeau, K. (2018, November 15). Hormone therapy for depression: Are the risks worth the benefits? Retrieved from https://www.health.harvard.edu/blog/hormone-therapy-for-depression-are-the-risks-worth-the-benefits-2018111615378.

[4] Huettel, S. A.; Song, A. W.; McCarthy, G. (2009), Functional Magnetic Resonance Imaging (2 ed.), Massachusetts: Sinauer, ISBN 978-0-87893-286-3

[5] Schnyer, D. M., Clasen, P. C., Gonzalez, C., Beevers, C. G. (2017). Evaluating the diagnostic utility of applying a machine learning algorithm to diffusion tensor MRI measures in individuals with major depressive disorder. Psychiatry Research: Neuroimaging, 264, 1–9. doi: 10.1016/j.pscychresns.2017.03.003

[6] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker: "Attention U-Net: Learning Where to Look for the Pancreas", 2018; arXiv:1804.03999.

[7] Fei Wang, Mengqing Jiang, Chen Qian, Shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang: "Residual Attention Network for Image Classification", 2017; arXiv:1704.06904.

[8] Masanari Kimura: "Intentional Attention Mask Transformation for Robust CNN Classification", 2019; arXiv:1905.02719.

[9] Wang, Z., Childress, A. R., Wang, J., Detre, J. A. (2007). Support vector machine learning-based fMRI data group analysis. NeuroImage, 36(4), 1139–1151. doi:10.1016/j.neuroimage.2007.03.072

[10] Karen Simonyan, Andrew Zisserman: "Very Deep Convolutional Networks for Large-Scale Image Recognition", 2015; arXiv: 1409.1556

[11]Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovic: "Going Deeper with Convolutions", 2014; arXiv: 1409.4842

[12] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, Zbigniew Wojna: Rethinking the Inception Architecture for Computer Vision, 2015; arXi: 1512.00567

[13]Lepping, R., Atchley, R., Chrysikou, E., Martin, L., Clair, A., Ingram, R. et al. (2016). Neural Processing of Emotional Musical and Nonmusical Stimuli in Depression. PLOS ONE, 11(9), e0163631. doi: 10.1371/journal.pone.0163631