CS230

Deep Learning for Classifying Heavy Drinkers from Normal Controls

Chuanqi Chen*
MACH Division
KLA-Tencor Corporation
chuanqi.chen@kla-tencor.com

Qingyu Zhao†
Department of Computer Science
Stanford University
qingyuz@stanford.edu

Abstract

There is speculation that adolescent brain is especially vulnerable to environmental insult such as heavy alcohol drinking. In this paper, we describe our pioneer work of using deep learning methods to classify heavy drinkers versus normal adolescents based on their 3D MRI images. In addition, we investigate importance of different brain regions in classifying heavy drinkers vs normal controls. We show that the result is convergent with existing literatures with respect to the brain regions that are more vulnerable to heavy drinking during adolescence, E.g., frontal and temporal lobes are more impacted than occipital lobe.

1 Introduction

The Human brain always changes through the life span, but the critical development is after birth throughout adolescence and into young adulthood. There is speculation that adolescent brain is especially vulnerable to environmental insult such as heavy alcohol drinking. Recent U.S. surveys estimate that 66% of 18-year-olds have drunk alcohol and about 25% report binge drinking. A deep understanding of if alcohol consumption can induce abnormal brain development and how different brain structures are impacted is important for advancing the prevention, prediction and control of the high risk adolescent alcohol-use-disorder (AUD).

Based on statistical group analysis, there are initial evidences showing that alcohol drinking might impact normal brain development during adolescence [1, 2]. One of our goals in this project is then to test whether neural-network-based methods can reasonably classify heavy drinkers versus normal adolescents based on their brain structural data. From the classification, we also want to identify patterns of the brain that contribute significantly to the classification as an indicator of AUD-related atropy, and we want to see if these patterns agree with existing evidences in the literature.

2 Related work

Existing analyses of AUD impact on brain structures are mainly based on statistical hypothesis testing. Pfefferbaum et al. [1] found that the cortical volume and thickness of grey matter, which are critical indicators of human cognitive abilities, exhibit deficit in the AUD group compared to the control group. Significantly impacted regions included frontal and temporal lobes of the brain. In a followup study [2], they continued to show that in the control group gray matter volume declined throughout adolescence and slowed in many regions in later adolescence, and youths who initiated

*chuanqi.chen@gmail.com— KLA Tencor

†Postdoc fellow, Stanford University (qingyuz@stanford.edu)— sponsor for this project.

CS230: Deep Learning, Winter 2018, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

heavy drinking exhibited an accelerated frontal cortical gray matter trajectory, divergent from the norm. Although these results have improved the understanding of AUD-related impact, the group analysis approach used in the above studies were performed on the entire cohort of subjects for inference. This strategy is known to have low reliability and generalizability; i.e., findings revealed in a specific cohort of subjects are not likely to be replicated on a new cohort.

Recent advances in machine learning [3], specifically deep learning [4, 5], have led to tremendous progress in disease analysis. In these studies, to identify critical regions impacted by a disease, a common practice is to first train a classification algorithm that can accurately differentiate healthy subjects from diseased ones based on their imaging data. Then measurements from the subset of regions highly influencing the classification outcome are identified as disease-specific biomarkers. Park et al. [6] built a linear classifier for the NCANDA dataset based on the subjects' low-dimensional cortical measurements (e.g., volume of each brain cortical region) derived from MRI data. The algorithm achieved 75% classification accuracy based on a "single" two-fold cross validation. However, Park et al. performed the analysis only on a small subset (34 extreme drinkers among the 134 drinkers) of the NCANDA dataset, and it is known that two-fold cross-validation exhibits high variation on small datasets such as theirs. Therefore in this project, we aim to perform a more rigorous classification analysis on the complete NCANDA dataset using 10-fold cross-validation. Moreover, we aim to develop deep learning methods directly operating on raw 3D images to reveal more spatially specific patterns (voxel-wise) rather than the coarse measurements used (region-wise) in their analysis. This could provide a more detailed understanding of the specific brain regions impacted by drinking.

3 Dataset and Features

National Consortium on Alcohol and NeuroDevelopment in Adolescence (NCANDA) magnetic resonance imaging (MRI) data is consisted of MRI images of total 808 adolescents, in which 674 adolescents meeting no/low alcohol or drug use criteria and 134 adolescents exceeding criteria. Some adolescents have up to 3 annual scans as a result of the follow-up scans of each subject. And NCANDA MR dataset(N808) has also manual measurements of different brain regions (aparc-left-hemisphere, aparc-right-hemisphere, aseg). These measurements are corrected for age, sex, scanner, race, brain size and socioeconomic status through a linear regression. There measurements data can be utilized in the future research work along with proposed deep learning methods with 3D MRI raw images.

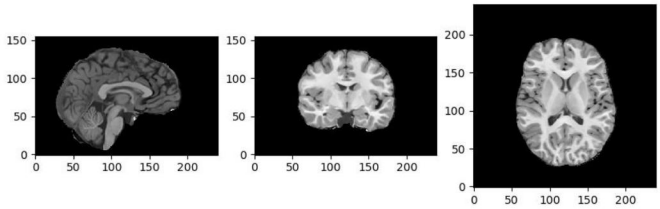

Figure 1: MRI 3D Image Slices along X, Y and Z Axes

Preprocessing of MRI image data was performed via the Biomedical Informatics Technology for Imaging Studies procedure, which includes denoising, bias field correction, and skull stripping by voting. The preprocessed images were then non-rigidly [affinely] registered to the sri24 atlas (Rohlfing et al. 2010). The registered images were also downsampled to 2mm spatial resolution.

4 Methods

Each subject's MRI data is three dimension image data, which were extracted from NCANDA MRI Dataset, read and cropped to 224x224x151 pixels using Nibabel Python library [7], thus there are 151 slices of 224x224 image data. Each slice of image are applied with Histogram-based intensity standardization.

The network is primarily based on SqueezeNet pre-trained model. Even though we started with bigger Network VGG19, but it didn't perform well due to limited dataset, however it might be beneficial for

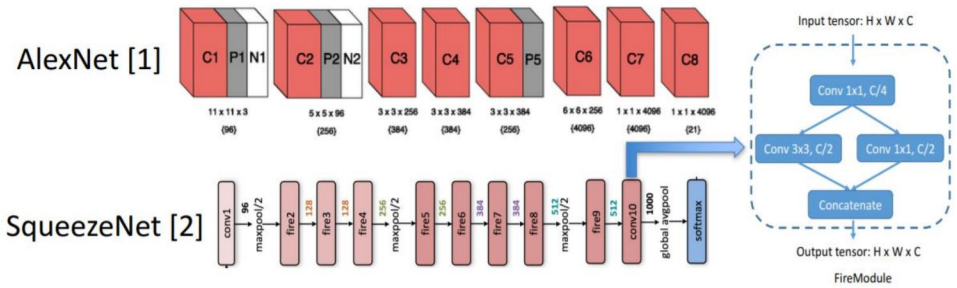

Figure 2: Pre-trained Model SqueezeNet vs. AlexNet [8]
Networkflow: SqueezeNet Feature Layers with earlier layers frozen -> Dropout-> Conv2d ->Dropout -> AdaptiveAvgPo0l3d -> Linear Classifier

future training should we have more MRI images of adolescents. Following research study [9] [10], We also compared between AlexNet and SqueezeNet and found out that SqeezeNet is much faster with even better accuracy for our experiment. Because we only have datasets of 808 subjects, we had to apply transferring learning techniques by freezing most of earlier layers and only unfreeze later layers of SqueezeNet.

Each MRI training image was scaled randomly with ratio between 0.9 and 1.1, shifted randomly between -10 and 10 pixels, rotated randomly between -10 and 10 degrees with 50% probability. Then 151 slices of MRI gray scale 224x224 were copied and converted to RGB format, resulting in data with dimension of 151x3x224x224.

Each 2-dimensional MRI image slice (224x224) was feed into SqueezeNet's feature extractor to obtain a 151x512x7x7 tensor, and after another convolution layer with the result of 151x512x7x7 tensor. A global average pooling layer was applied to reduce these features to 151x512, and then was further reduced to 512-dimensional vector with max pooling across 151 slices. Then after a linear classifier, sigmoid actuation function was used to predict results.

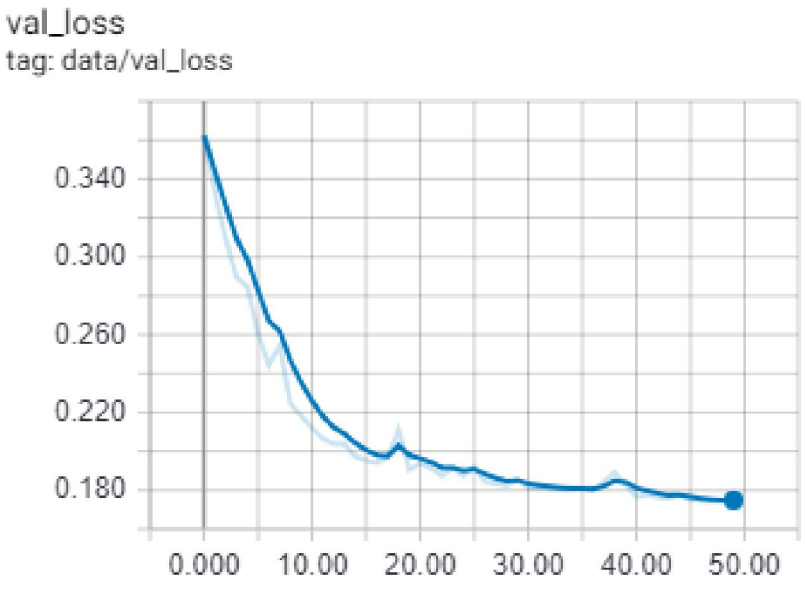

Figure 3: Validation Loss

The datasets were split into roughly 75% for training, 15% for validation and 15% for testing. We optimized the model in order to reduce validation loss with binary cross-entropy. The model trained with training sets were used to evaluate validation set to choose the best model if the validation loss was improved.

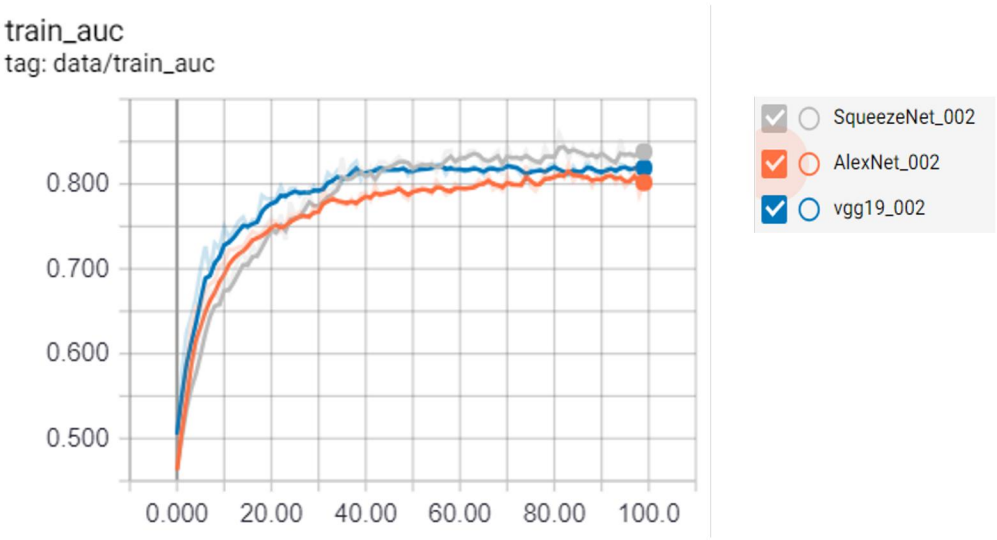

Figure 4: Training Accuracy

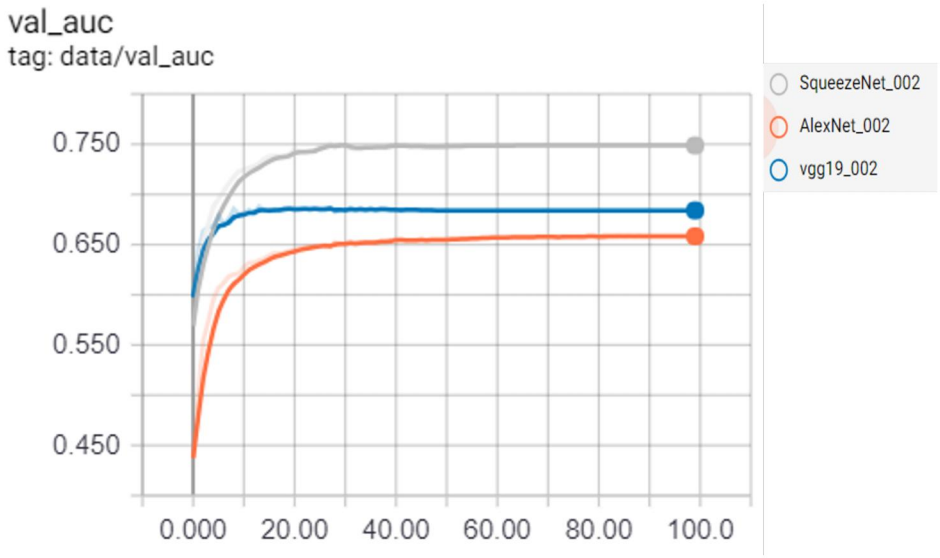

Figure 5: Validation Accuracy

5 Experiments/Results/Discussion

All hyper parameters are programmed as command line arguments, which enables us to use script to loop through the ranges of these hyper parameters by running experiments to find the best settings which minimize validation loss and thus increase the accuracy.

For example, we looped through three options of pretrained models (SqueezeNet, AlexNet and VGG19) through same experiments and compared the validation loss and prediction accuracy and also precision-recall metrics and then choose the best model, SqueezeNet, Refer to Fig. 4 Fig: 5,

With our deep learning methods, fully end to end prediction without any manual work, we are able to achieve 74.87% prediction accuracy, on par with existing machine learning methods applying on manual measurements from MRI images.

6 Conclusion/Future Work

Pre-trained Model SqueezeNet performs the best

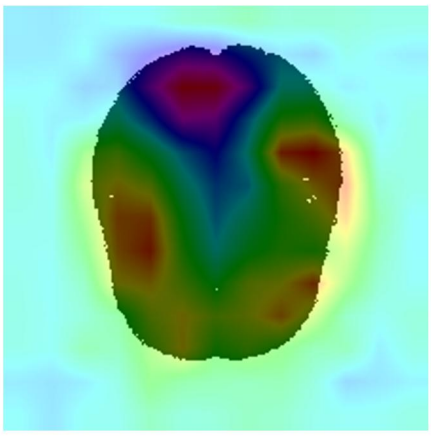

Figure 6: Heatmap

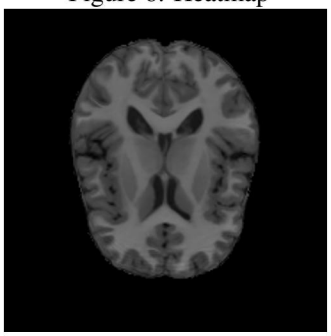

Figure 7: MRI Image slice at Z=0

Pre-trained model Vgg19 is performs second due to limited dataset of 808 MRI image sets. It might be good candidate model should we have gathered more dataset in the future.

Heat Map shows interesting results which regions of brain playing more important roles in classifying the heavy drinkers. And it might provide insight of what regions of brain of adolescence are more vulnerable and impacted by heavy drinking.

7 Code Implementation

The code are located in github: `https://github.com/chuanqichen/volumenet`

8 Contributions

I am the only team member of this project. My sponsor, Qingyu Zhao, provides tremendous domain knowledge and also provides me with precious MRI image set. His validation and confirmation of the project results are very important for me to achieve the realistic goal. And my mentor, Ahmadreza Momeni, and many TA such as Steven provide great deal of helps and guidance such as how to choose pretrained network, how to deal with model overfit and how to implement heatmap, etc.

References

[1] A. Pfefferbaum, T. Rohlfing, K. M. Pohl, B. Lane, W. Chu, D. Kwon, et al. Adolescent development of cortical and white matter structure in the ncanda sample: Role of sex, ethnicity, puberty, and alcohol drinking. *Cereb Cortex.*, 26(10):4101–4121, 2016. doi: 10.1093/cercor/bhv205.

[2] Adolf Pfefferbaum, Dongjin Kwon, et al. Altered brain developmental trajectories in adolescents after initiating drinking. *American Journal of Psychiatry.*, 175(4):370–380, 2017.

[3] Miles N. Wernick, Yongyi Yang, Jovan G. Brankov, Grigori Yourganov, and Stephen C. Strother. Machine learning in medical imaging. *IEEE Signal Process Mag.*, 27(4):25–38, 2010.

[4] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, 19:221–248, 2017.

[5] Eli Gibson, Wenqi Li, Carole Sudre, Lucas Fidon, et al. Niftynet: a deep-learning platform for medical imaging. *Computer Methods and Programs in Biomedicine*, 158:113–122, 2018.

[6] Sang Hyun Park, Yong Zhang, Dongjin Kwon, Qingyu Zhao, et al. Alcohol use effects on adolescent brain development revealed by simultaneously removing confounding factors, identifying morphometric patterns, and classifying individuals. *Nature Scientific Report*, 8(8297), 2018.

[7] Neuroimaging python library. http://nipy.org/nibabel/. Read / write access to some common neuroimaging file formats.

[8] The impact of squeezenet. https://aspire.eecs.berkeley.edu/wiki/_media/eop/2017/bichen_talk_slides.pdf.

[9] Ilya Sutskever Krizhevsky, Alex and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 2012.

[10] Forrest N. Iandola. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and< 1mb model size. *arXiv*, 2016.