Machine Learning Aided Prediction of Family History of Depression

Allen Liu
Rocky Point High School
Rocky Point, USA
allenliu1136@hotmail.com

Bryant Liu
Rocky Point High School
Rocky Point, USA
bryantliu11235@gmail.com

Daniel Lee
Commack High School
Commack, USA
dani.cmh.lee@gmail.com

Myrna Weissman
Department of Psychiatry
Columbia University
New York, USA
weissman@nyspi.columbia.edu

Jonathan Posner
Department of Psychiatry
Columbia Unversity
New York, USA
posnerj@nyspi.columbia.edu

Jiook Cha
Department of Psychiatry
Columbia University
New York, USA
jiook.cha@nyspi.columbia.edu

Shinjae Yoo
Computer Science Initiative
Brookhaven National Lab
Upton, USA
sjyoo@bnl.gov

***Abstract*—Increased risk for psychopathology in the offspring of depressed parents has been widely known. The brain may mediate the effects of risk of familial depression on the offspring via shared genetic and environmental factors. Conventional brain imaging studies to test this mediation effects primarily use *a priori* knowledge to select a subset of brain imaging-derived features. Despite of the existing positive results supporting the notion of the brain as an endophenotype for familial depression, no quantitative assessment has been performed regarding to what extents the complex brain structure contains information about familial depression. To this end, here we aim to predict whether an individual has a history of familial depression. We propose a data-driven, unbiased, and rigorous machine learning approach using multimodal brain features (e.g., grey matter morphometry based on T1-weighted images and structural connectome based on probabilistic diffusion tractography) to capture the complex representations of brain structure. We implemented logistic regression (LR) with regularization, support vector machine (SVR), and graph convolutional neural network (GCN). Our models show promising cross-validated classification accuracy: 97.78% (LR), 93.67% (SVM) and 89.58% (GCN). Brain features with greatest weights in the models include brain regions previously implicated in the depression literature (e.g., emotion regulation frontal-limbic circuit) as well as new regions. Results suggest a large impact of familial depression on the brain structure and connectome. Results also highlight potentials for prediction of risk for psychopathology in a data-driven fashion using cost-effective, simple, ubiquitous brain images.**

Keywords—mental disorder, familial depression, machine learning, graph convolutional network

I. Introduction

Mental disorder is a common human illness. Studies have revealed that approximately 44 million or 19 % of American adults face mental illness each year. Among them, 16 million people (7% of Americans) have suffered at least one major depressive episode in the past year [1]. Previous studies have shown that children born to the parents with depression have a higher risk for psychopathology such as depression, ADHD, and drug abuse [2].

The effects of the higher risk on the offspring are mostly likely to be mediated by the brain. Prior studies assessing a subset of the brain, informed by *a priori* knowledge, demonstrate significant associations of familial depression, brain features, and clinical/behavioral measures [4][6]. An important remaining question is to what extent familial depression exerts an effect on an individual brain. Quantifying this effect may contribute to predictive modeling of risk for psychopathologies in a given individual. To increase reproducibility and generalizability of a model, this may be conducted in an unbiased, data-driven manner.

To this end, here we test how confidently a model, using a wide array of brain features representing brain morphology and structural connectome, could predict the existence of family history of depression in a given individual. Key to this goal is a rigorous analytic method to handle large feature space compared with a relatively small number of observations. Along with the development of computational science, machine learning algorithms have been proven powerful tools for processing high-dimensional datasets and precisely classifying disease features thus helping in clinical diagnosis. A recent study showed that deep neural network model trained on skin images predicted skin cancers with a similar precision to that of dermatologists' [3]. Analogous to this example, it is feasible a machine learning model trained on neuroimaging-derived features to predict whether an individual is at risk for recurring psychopathology issues [4].

Convolutional Neural Network is a category of neural networks in deep learning algorithms which have been proven to be powerful in classification problems in image, video and

978-1-5386-3161-4/17/$31.00 ©2017 IEEE

sound data [5]. They offer efficient extraction of relevant patterns in large and high-dimensional datasets. The basic architecture of a Convolutional Neural Network includes a convolutional layer, activation layer, pooling layer, and fully connected or classification layer.

Although CNNs are extremely useful for these tasks, they may not be easily applied to non-Euclidean domains, such as data structured as graphs, because the convolution and pooling operators cannot be applied to irregularly structured data. These problems were solved by replacing the convolution operator with a Fourier transform, and replacing the pooling operator with graph coarsening methods [5]. Since graphs offer a versatile structure to model other forms of data including social networks, biological networks, and infrastructure networks, Graph Convolutional Networks (GCNs) expand the limitation of input data in regular CNNs, and allows us to extract important information from irregular/graph structured data. Recent experimentation of GCNs have been applied to image data, such as MNIST [5], but not to brain imaging analysis.

Here we test machine learning models to classify an individual with a history of familial depression vs. without one using brain structure and connectome data. Using this approach, we additionally aim to discover the brain features associated with familial depression in a hypothesis-free way.

There are several challenges present in application of machine learning to the brain imaging data: 1) Hyperparameter tuning, and implementation of cross validation on the data to avoid over fitting. 2) Feature selection and data processing to achieve the best performance from the models. 3) Discovering the link between brain regions and the risk of psychological problems in each individual.

II. Materials and Methods

A. Data

Risk level of each participant was defined based on the first generation (G1). Offspring of G1 (Generations 2 and 3, G2 and G3) was given a high-risk label if the G1 parent had a history of major depressive disorder (MDD) [5]. For the current study, the MRIs of 93 participants from the G2 and G3 samples were collected and obtained, and each participant was labeled as either 0—low or 1—high-risk. A more detailed explanation of the family depression cohort can be found elsewhere [2].

B. Methods

To extract brain imaging-derived features, we analyzed two different MRI: T1-weighted structural images and diffusion weighted images. From structural images, we obtained brain-regional morphometric information, such as thickness, area, and volume, by performing automated brain segmentation and parcellation using Freesurfer image analysis suite v6.0[1]. From diffusion images, we obtained white matter connectivity matrices using probabilistic tractography and an algorithm to filter false positive tract estimates; detailed description about the method and codes are available elsewhere[2]. These connectivity matrices were weighted using two different methods: the number of tracts estimates and their mean length. The brain features were then normalized first to keep data at a comparable range, and second to keep features weighted equally. Standardized features also generally improve efficiency of machine learning algorithms.

Due to the small sample size, we used a stratified K-Fold cross validation to split the data into training, validation, and testing sets. Stratification rearranges the data to ensure each fold is representative of samples of each class, which reduces variance of cross validation estimation in the models. In this experiment, it was rearranged to have equal numbers of low and high-risk individuals. Every model used a stratified 10-Fold cross validation. For GCN, we used a validation set to tune hyper-parameters. However, both logistic regression and SVM may be tuned for the regularization parameter on a validation set but we did not use the validation data as we used the default regularization parameter.

Since the logistic regression and SVM are not as prone to overfitting as neural networks, only a train and test set were needed, however the datasets were still created through K-Fold cross validation. Each dataset was rotated every time the algorithms were trained. Additionally, the models were all repeated ten times with stratified 10-fold cross validation in order to account for fluctuations in accuracy due to the random initialization of weights in each model. The average accuracy and standard deviation of the ten runs were calculated for each model.

A final Receiver Operating Characteristic (ROC) curve was drawn for each of the models for comparison, and the area under the curve was calculated to assess the prediction strengths. The ROC curves show the relationship between sensitivity and specificity of a model. Furthermore, since we are using a logistic regression, feature weights were easily extracted, and then interpreted. For instance, the highest weight or the weight histogram distribution impact accuracy of the prediction.

III. Results

A. Feature Selection

Each participant had over 15,000 brain features, where all zero and duplicate features were removed through data preprocessing which reduced the total number of features to 7876 usable features. The features were then sorted based upon the correlation between each feature and risk of depression. To normalize the data, a z-score was taken to ensure that input into the machine learning models would be efficient and accurate. Various numbers of features were selected from the sorted list, which were then tested to determine the optimal amount of data for the models. In our dataset, the peak for the linear SVM and the logistic regression was around 900 and performance started to decay rapidly after 3200 features. Therefore, the models in this experiment were all fitted

[1] https://surfer.nmr.mgh.harvard.edu/

[2] https://academiccommons.columbia.edu/catalog/ac:n5tb2rbp1f

with the top 900 features for each individual for the best possible performance in prediction accuracy.

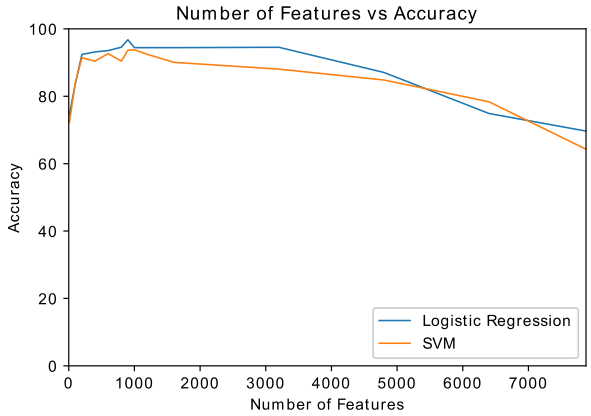

Fig 1. Plot of number of features vs accuracy for the logistic regression and SVM models. GCN was omitted due to the high computational cost and the extensive amounts of time it takes to train and test.

B. Training

Training, validation, and testing datasets were created through stratified 10-fold cross validation. Due to the small sample size, each fold contained between only 8-10 samples each, which had to be rotated as a part of cross-valuation. Consequently, the GCN, which initializes weights randomly and has more tuning parameters, had fluctuations in performance accuracy, and was unstable. For evaluation purpose, we repeated ten-fold cross validation and ten times. The result mean accuracy and its standard deviation are reported in Table I.

Table I. Prediction Accuracy Stability

Model	*Average %*	*Standard Deviation*
Logistic Regression	96.78	0.00
SVM	93.67	1.11e-16
GCN	89.58	2.03

C. Reciever Operating Characteristic

ROC curves are designed to depict the tradeoff between sensitivity and specificity, so curves were constructed for each model to visualize the predictive power. The diagonal dashed line in Fig. 2 shows the performance of binary classification based upon random chance. In our case, results were impressive with Logistic Regression having 99% area under the curve (AUC), which leaves very little for improvement. The SVM with an AUC of 98% and the GCN with an AUC of 96% also yielded promising results.

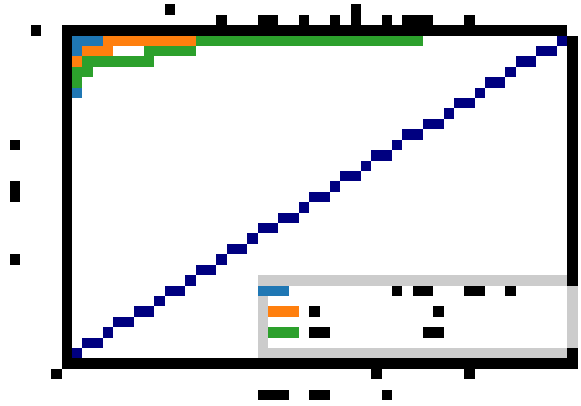

Fig. 2. ROC curve comparison of each model, and their respective area under the curve.

D. Feature Weight Extraction

To aid neuroscientific interpretation, we extracted weights of the features from the logistic regression model (Fig. 3).

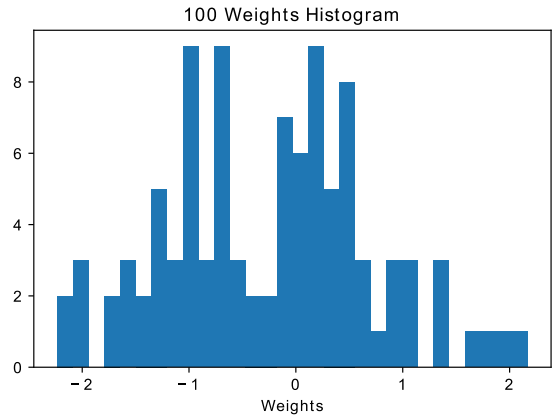

Fig 3. Histogram of the weight distributions.

We observed redundancy in features. With top 100 features selected, most significant feature weights were greater than 2. On the other hand, with 1000 features selected, significant feature weights dropped as low as 0.2; nevertheless, the prediction accuracies of both models stayed about the same level. This suggests that there are large number of redundant features and weight drifting. Indeed, in our brain imaging derived measures, we used different ways to weight structural connectome, such as tract counts, mean length, as well as different morphometric measures, such as surface area, volumes, and thickness.

The top ten features of the model (Table II) showed the brains regions and connections that have been previously implicated in emotion regulation and decision making, such as the amygdala, orbitofrontal gyrus, entorhinal cortex, and insula. Also, brain regions that were not implicated in the depression literature also appeared, such as occipital and supramarginal gyrus. This shows potential utility of this approach as a hypothesis-generating tool in neuroimaging.

Table II. Ranked Feature Weights (top 10)

Feature	*Weight Values*
Mean length between rh-inferiortemporal and lh-entorhinal cortex	0.274
Mean length between rh-lateraloccipital and lh-inferiortemporal gyrus.	-0.265
Mean length between rh-parsorbitalis and lh-precuneus	0.263
Tract counts between lh-Amygdala and lh-parahippocampal gyrus	0.251
Mean length between rh-lateraloccipital and lh-supramarginal gyrus	0.250
Mean length between lh-insula and ctx-lh-supramarginal gyrus	0.250
Tract counts between lh-superiorfrontal and lh-precentral gyrus	0.250
Mean length between lh-postcentral and lh-medialorbitofrontal gyrus	-0.246
Tract counts between rh-pericalcarine and rh-caudalmiddlefrontal gyrus	-0.239
Mean length between rh-superiortemporal and lh-entorhinal cortex	0.239

IV. Conclusion

Here we demonstrate that machine learning models based on large-scale features estimated across the whole brain and the whole connectome can reliably predict existence of familial depression in a completely data-driven fashion. The major contributions of this study to the field are: 1) Applying multi-dimensional brain imaging data to classify familial depression, an important risk factor for psychopathology; 2) Maximizing and testing prediction accuracy using a rigorous cross validation scheme; 3) Discovering brain features associated with the risk factor for psychopathology to inform future research.

The high (cross-validated) classification accuracy (96.78%) shows the significant impact of familial depression on the brain structure and connectome. This study may inform development of a clinically-useful model to predict risk for psychopathophysiology using brain imaging-derived phenotypes. For example, it may be interesting to test whether this familial depression model could inform a model predicting maladaptive emotion or cognition (e.g., impulsivity or anhedonia) as a multi-task learning model. Also, this study shows machine learning application to brain imaging data can be used as a rigorous exploratory, hypothesis-generating tool, to overcome a biased, under-powered analytic method.

Acknowledgment

This work was supported in part by BNL LDRD 16-039, DE-SC0012704, and NIMH R01-MH036197. We would like to thank Brookhaven National Lab for providing us the facility for working on this project, and to the High School Research Program that provided us the opportunity to have on-site access to BNL and learn from real experience.

References

[1] http://www.nimh.nih.gov/health/statistics/prevalence/any-mental-illness-ami-among-adults.shtml

[2] Weissman MM, Wickramaratne P, Gameroff MJ, Warner V, Pilowsky D, Kohad RG, Verdeli H, Skipper J, Talati A, "Offspring of Depressed Parents: 30 Years Later," in Am J Psychiatry, vol. 173, pp. 1024-1032, October 2016.

[3] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S, "Dermatologist-level classification of skin cancer with deep neural networks," in Nature, vol. 542, issue 7639, pp. 115-118, Febuary 2017.

[4] Cha J, Guffanti G, Gingrich J, Talati A, Wickramaratne P, Weissman MM, and Posner J, "Effects of Serotonin Transporter Gene Variation on Impulsivity Mediated by Default Mode Network: A Family Study of Depression," in Cerebral Cortex, pp. 1-11, 201

[5] Defferrard M, Bresson X, and Pierre V, "Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering," in CoRR, vol. abs/1606.09375. https://arxiv.org/abs/1606.09375

[6] Posner, J., Cha, J., Wang, Z., Talati, A., Warner, V., Gerber, A., Peterson B., Weissman, M. (2016). Increased Default Mode Network Connectivity in Individuals at High Familial Risk for Depression. *Neuropsychopharmacology*, *41*(7), 1759–1767. https://doi.org/10.1038/npp.2015.342