Autism Diagnosis Using Transformer Architecture on Resting-State f-MRI

Connor B. Parish[a]

[a] cbp5nd@virginia.edu, 1207 Wertland St. Charlottesville, Va, 716-949-2354

Number of words: 1,958

Number of figures and tables: 2

Number of equations: 0

Number of supplements: 0

Number of references: 20

Autism Diagnosis Using Transformer Architecture on Resting-State f-MRI

Connor B. Parish[a]

[a] cbp5nd@virginia.edu, 1207 Wertland St. Charlottesville, Va, 716-949-2354

Abstract

Autism is a developmental disorder characterized by challenges with social skills, repetitive behaviors, speech, and nonverbal communication. The main method of diagnosis for autism is based off of a doctor's analysis of an individual patient's history and behaviors. Research is being done to try to use fMRI scans to diagnosis autism in order to improve diagnosis consistency and provide more information to improve treatment choices. The goal of this research is to gauge the efficacy of using a Transformer architecture to improve the automated diagnosis performance on the ABIDE I and ABIDE II datasets through the use of transfer and multi-task learning. The final model trained using a multi-task training method was able to achieve an accuracy of 0.689% on a subset of the ABIDE datasets. The most comparable previous work achieved an accuracy of 0.652% on the same subset of subjects.

Keywords: Autism, ABIDE, Transformer, Self-Attention, Machine Learning, Transfer Learning, Multi-task

Introduction

The main method of diagnosis for autism is based off of a doctor's analysis of an individual patient's history and behaviors. Research is being done to try to use fMRI scans to diagnosis autism in order to improve diagnosis consistency and provide more information to improve treatment choices (Dekhil et al., 2019; Dvornek et al., 2017; Heinsfeld et al., 2017; Hull et al., 2017; Lau et al., 2019; Thomas et al., 2020). The goal of this research is to improve upon current fMRI autism diagnosis techniques to increase the accuracy of the diagnosis towards the ultimate goal of reaching clinical capable consistency.

The Autism Brain Imaging Data Exchange (ABIDE) is an aggregation of functional and structural brain imaging data collected from around the world with the goal of accelerating the neural bases of autism. This exchange consists of two datasets, ABIDE I and ABIDE II, that combined consist of brain imaging data from 1,060 patients with ASD and 1,166 control patients. ABIDE I has preprocessed datasets publicly available, but ABIDE II requires preprocessing before research can be done on it. Most research utilizing the ABIDE datasets has been focused on using structural brain imaging data, but as computational tools have improved, more focus has turned to analyzing the functional data. One way to utilize the functional data is through region of interest activation extraction. This method maps voxels within the 3D MRI scan at each time step to regions within the brain. Then, the activation of each region is extracted using this atlas mapping at each of the f-MRI time steps. The resulting data is a matrix that is the number of regions of interests in the atlas by the number of time steps of the scan.

The most relevant previous work to this project is *Identifying Autism from Resting-State fMRI Using Long Short-Term Memory Networks* (Dvornek et al., 2017). In this work, the authors used LSTM architectures to classify ASD on roi data from the ABIDE I preprocessed dataset. One significant difference between this work and my project is that the authors used interpolation to accommodate MRI scans with varying time-steps. This allowed them to use the entire ABIDE I preprocessed dataset, while I only used scans from sites that used time-steps of 2s. Thus, the authors trained and evaluated their models on 539 individuals with ASD and 573 control individuals from 17 international sites. The authors were able to achieve a classification accuracy of 68.5%.

A Transformer is a novel Machine Learning Architecture first proposed in *Attention Is All You Need* (Vaswani et al., 2017). This architecture utilizes a concept called self-attention. Through self-attention, the Transformer is able to draw inferences on how different features within a given input interact with each other. Through this, the model can learn interactions across entire inputs of varying size, which was previously impossible without very large feature extractors, such as convolutional kernels. Transformers have revolutionized multiple fields of machine learning, most notable Natural Language Processing. Natural Language Processing is the use of computational tools to draw inference from human language. The BERT architecture, which utilizes a Bi-Directional Transformer, greatly improved NLP tasks ranging from classification to question answering.

One significant contributor to the success of the Transformer is its ability to utilize Transfer learning techniques on a significant scale. Transfer learning is where a model developed for a task is reused as the starting point for a model on a different task. The BERT model takes advantage of transfer learning through the use of pre-training tasks. Most commonly, the model is pre-trained using the parallel tasks of next-sentence prediction and masked work prediction. The next-sentence prediction is a binary classification task in which the model predicts if a pair of sentences were written in order. Masked word prediction is the task in which around 15% of the words of a sentence are hidden from the model, and the model learns to predict the probability that the masked words are a given word within the training dictionary. This parallel training of multiple downstream tasks is called multi-task learning. This pre-training is performed on large textual datasets, such as Wikipedia. Through this pre-training, the Transformer is able to learn how to draw meaningful insights from textual data. Then, the pre-trained Transformer is used for downstream tasks that may not have large datasets. Therefore, the models are able to perform well, generalize well, and avoid overfitting despite being trained on relatively small datasets.

The goal of this project was to utilize the Transformer architecture to draw generalizable insights from resting-state f-MRI roi timeseries data. This would be done through multi-task pre-training on a large dataset of f-MRI scans and then applying the trained weights to downstream tasks with smaller datasets. One major issue within the field of modeling the human connectome through resting-state f-MRI activation data is the heterogeneity of the data. This heterogeneity is a result of many factors including varying MRI parameters, imaging artifacts, varying human subject populations, and highly varying preprocessing pipelines. Although this level of heterogeneity makes drawing meaningful inferences with shallower deep learning models nearly impossible, Transformer's ability to overcome this hurdle is one major contributor to their success in multiple varying machine learning tasks.

<u>Results</u>

The models were trained and evaluated on subjects from both ABIDE I and ABIDE II that came from scanning sites with time-steps of 2s that passed the preprocessing quality checks. This resulted in 464 subjects with ASD and 485 control subjects. The data was split up into training, validation, and test sets with proportions of .85, .05, and .1 respectively.

The first step of the training processes was pre-training. In the first attempt, pre-training was performed using the ABIDE, Addiction Connectome Preprocessed Initiative, and ADHD200 datasets. Once filtered, the ACPI and ADHD200 datasets contributed 155 and 82 additional subjects respectively. However, due to a lack of funds, previously preprocessed versions of these datasets were used. Although the atlas was the same, the preprocessing pipelines varied from the ABIDE I preprocessed and my pipelines greatly.

The pretraining tasks were gender prediction, age prediction, and next-brain state prediction. During next-brain state prediction, the model was trained to create a prediction of the next activation state after the extracted timesteps. Unfortunately, the model was unable to learn properly on the pre-training tasks. The only task with reasonable performance was gender prediction. This was likely due to the heterogeneity of the ACPI and ADHD200 datasets. The model was able to achieve better pre-training with just the ABIDE datasets. As well, the model was able to perform well on the next-activation state prediction when a sequence length of 10 was used, but it performed poorly on the sequence length of 90. The sequence length of 90 was maintained due to the better performance of the other tasks, and the next-brain state prediction task was dropped.

Since there was no additional data to pre-train on, it was decided to perform the training using multi-task training. Thus, the model was trained using the parallel tasks of ASD classification, gender prediction, and age prediction. The model was also trained with and without positional embeddings, but the model consistently performed better with position embeddings so they were included in the multi-task training. The multi-task trained Transformer model performed better than the LSTM or single-task Transformer models as seen in Table 1.

Task	Set	LSTM	Transformer	Transformer with position	Mulit-task training
Asd_acc	Train	0.899	0.989	0.992	0.974
	Test	0.652	0.618	0.639	0.689
Gender_acc	Train	na	na	na	0.998
	Test	na	na	na	0.544

Table. 1: Results from Autism Spectrum Disorder Classification on ABIDE I and ABIDE II Datasets.

Discussion

The first major thing to note is the context of the training and evaluation of the models. The models were trained on a free version of Google Colab notebooks. Thus, the amount of RAM and GPU computing time was limited. For this reason, the training was not done using cross-validation, which is used to get mean and standard deviation values for the model performance over multiple permutations of dataset distributions. The next step would be to rerun these experiments with cross-validation to further verify the results.

The main advantage of transfer learning comes from the ability of the model to be trained on a larger dataset than the dataset of interest. Due to a lack of resources and publicly available preprocessed datasets, this goal was not achieved. In order to improve model performance and reduce overfitting, more f-MRI datasets should be preprocessed using the CPAC pipeline. Then the Transformer model should be pre-trained on this data using the gender classification, age prediction, and next-brain state prediction. With enough data these tasks could greatly improve the Transformer's ability to draw generalizable insights from the connectome data. Another option would be pre-training using a masked roi task. In this task, random rois within the input would be hidden from the model, and the model would be trained to predict the hidden timesteps.

Another way to increase the amount of data would be to create a model able to handle varying time-steps. Although

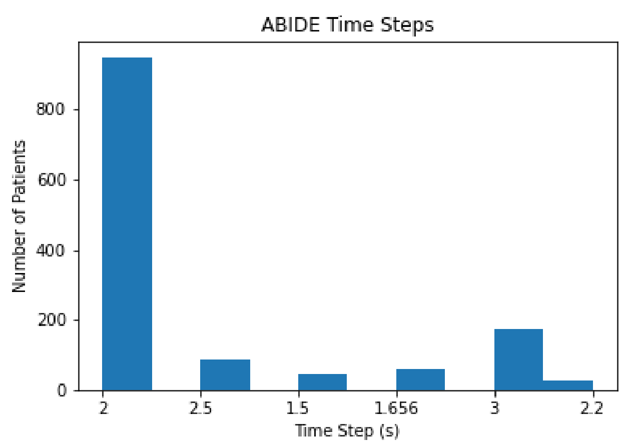

Fig. 1: Distribution of time steps within the ABIDE datasets.

most scanning sites included in ABIDE used time steps of 2s (Fig 1), this is not always the case for publicly available datasets. For example, the Healthy Brain Network dataset, which is the largest preprocessed dataset I found, did not have any data from sites with time steps of 2s. Most previous studies utilizing the ABIDE I preprocessed dataset overcame this issue through interpolation of the data to 2s time steps. However, this interpolation fundamentally decreases the information within the data and may decrease the performance of the model. For this reason, I decided to not use interpolation in these experiments. Future research could be done on how including interpolated data impacts the performance of the model. Another alternative would be to include the time step as a feature input into the model. This would allow the model to change the way it processes the roi activation time steps depending on the time steps.

One way to do this would be to have different positional encodings for different time steps. Position encodings are the way that the model learns how to interpret the distance between inputs. Thus, having unique encodings for each time step could allow the model to appropriately adapt. However, this could also be another source of overfitting if some time steps do not have enough data or different time step categories had varying subject population distributions.

Another interesting experiment would be to analyze the performance of the model on varying sets of heterogenous data. For example, the model could be evaluated on a dataset preprocessed with a different pipeline than the one it was trained on. With enough data, it would be possible that the Transformer could learn to handle a variety of heterogenous data. This would have the major benefit of allowing the model to be pre-trained on a larger dataset of already available preprocessed datasets. As well, data from a single subject could be preprocessed in multiple ways to produce more data. This would be similar to the use of data augmentation for increasing the effective size of a dataset.

Materials and Methods

The first major step of the project was choosing the preprocessing pipeline. Since the ABIDE dataset is the main data source for the project, a pipeline that was used to create the ABIDE I preprocessed dataset was to be chosen. The Configurable Pipeline for the Analysis of Connectomes (CPAC) was chosen due to its widespread use, especially within projects with similar aims as this Capstone.

The next major step was to choose the parameters of the preprocessing pipeline. Although the majority of the parameters have widely used standards, the major decisions are whether to include global signal regression and/or band-pass filtering. "Global signal regression (GSR) is one of the most debated preprocessing strategies for resting-state functional MRI. GSR effectively removes global artifacts driven by motion and respiration, but also discards globally distributed neural information and introduces negative correlations between certain brain regions" (Li et al., 2019). Band-pass filtering is used to remove frequencies of the raw fMRI data that are not of interest. Thus, the goal is to improve the signal to noise ratio. It was decided that the pipeline would not utilize global signal regression but would use a standard fMRI band-pass filter (0.01 – 0.1 Hz). The main driver of this decision was that these parameters were chosen by the majority of projects utilizing ABIDE I preprocessed to derive insight from resting-state connectomes.

With the pipeline chosen, the next step was to implement the preprocessing pipeline. I first implemented the pipeline on an old computer that I have, but it did not have enough computational power. Thus, I implemented the pipeline using AWS EC2 instances and an S3 bucket. To verify the pipeline, subjects in ABIDE I were preprocessed and the regions of interest data were compared to the publicly available ABIDE I preprocessed. Next, the ABIDE II scan sites that used time steps of 2s were preprocessed. This cost around $110 when taking into account computational resources, storage, and data transfer cost. On average the pipeline cost around 20 cents per subject.

Next, the subjects were split into train, validation, and test sets. This was done on the subject level to ensure that the validation and testing sets were truly independent from the training set. The next step was extracting even lengthened segments from the time series data. Although Transformers can handle inputs of varying size, this was done to increase the total number of inputs. Following the precedent set by *Identifying Autism from Resting-State fMRI Using Long Short-Term Memory Networks* (Dvornek et al., 2017), an average of 10 segments of 90 time steps each were extracted from each subject.

The models were implemented using Google's Keras deep learning python library. The LSTM used was chosen beset model from the Dvornek et al. paper. The best-chosen Transformer model had 8 attention heads and a hidden layer size of 32. A dropout rate of 0.5 was implemented between each of the downstream layers. Three classification tokens were used to achieve the three classification tasks in parallel. The gender classification and age prediction tokens were fed into a 20 noded dense neural net layer with relu activation. The gender dense layer was fed into a sigmoid activated single noded layer, and the age prediction layer fed into a linear activated single noded layer. The ASD classification neural net was identical to the gender classification one. The next-brain state prediction downstream layers were two Dense layers with 200 nodes each. The first was relu activated, and the second was linearly activated. Since the next-brain state prediction neural net contained significantly more weights than the other task specific layers, it is likely that the lack of data impacted its ability to perform more significantly.

Training occurred on a free tier Google Colab notebook with a GPU. Cross-validation training was attempted, but it was not possible due to a lack of RAM. All scripts used to perform the experiments listed above are publicly available at https://github.com/Connorpar/asd_diagnosis_fmri .

References

Bachman, P., Hjelm, R. D., & Buchwalter, W. (2019). Learning Representations by Maximizing Mutual Information Across Views. *ArXiv:1906.00910 [Cs, Stat]*. http://arxiv.org/abs/1906.00910

Baevski, A., & Auli, M. (2019). Adaptive Input Representations for Neural Language Modeling. *ArXiv:1809.10853 [Cs]*. http://arxiv.org/abs/1809.10853

CDC. (2020a, February 11). *Healthcare Providers | Autism Spectrum Disorder (ASD) | NCBDDD | CDC*. Centers for Disease Control and Prevention. https://www.cdc.gov/ncbddd/autism/hcp-screening.html

CDC. (2020b, March 26). *Autism and Developmental Disabilities Monitoring (ADDM) Network | CDC*. Centers for Disease Control and Prevention. https://www.cdc.gov/ncbddd/autism/addm.html

Dekhil, O., Ali, M., El-Nakieb, Y., Shalaby, A., Soliman, A., Switala, A., Mahmoud, A., Ghazal, M., Hajjdiab, H., Casanova, M. F., Elmaghraby, A., Keynton, R., El-Baz, A., & Barnes, G. (2019). A Personalized Autism Diagnosis CAD System Using a Fusion of Structural MRI and Resting-State Functional MRI Data. *Frontiers in Psychiatry*, *10*. https://doi.org/10.3389/fpsyt.2019.00392

Dvornek, N. C., Ventola, P., Pelphrey, K. A., & Duncan, J. S. (2017). Identifying Autism from Resting-State fMRI Using Long Short-Term Memory Networks. *Machine Learning in Medical Imaging. MLMI (Workshop)*, *10541*, 362–370. https://doi.org/10.1007/978-3-319-67389-9_42

Fan, L., Su, J., Qin, J., Hu, D., & Shen, H. (2020). A Deep Network Model on Dynamic Functional Connectivity With Applications to Gender Classification and Intelligence Prediction. *Frontiers in Neuroscience*, *14*. https://doi.org/10.3389/fnins.2020.00881

Heinsfeld, A. S., Franco, A. R., Craddock, R. C., Buchweitz, A., & Meneguzzi, F. (2017a). Identification of autism spectrum disorder using deep learning and the ABIDE dataset. *NeuroImage : Clinical*, *17*, 16–23. https://doi.org/10.1016/j.nicl.2017.08.017

Heinsfeld, A. S., Franco, A. R., Craddock, R. C., Buchweitz, A., & Meneguzzi, F. (2017b). Identification of autism spectrum disorder using deep learning and the ABIDE dataset. *NeuroImage : Clinical*, *17*, 16–23. https://doi.org/10.1016/j.nicl.2017.08.017

Hull, J. V., Dokovna, L. B., Jacokes, Z. J., Torgerson, C. M., Irimia, A., & Van Horn, J. D. (2017). Resting-State Functional Connectivity in Autism Spectrum Disorders: A Review. *Frontiers in Psychiatry*, *7*. https://doi.org/10.3389/fpsyt.2016.00205

Khodatars, M., Shoeibi, A., Ghassemi, N., Jafari, M., Khadem, A., Sadeghi, D., Moridian, P., Hussain, S., Alizadehsani, R., Zare, A., Khosravi, A., Nahavandi, S., Acharya, U. R., & Berk, M. (2020). Deep Learning for Neuroimaging-based Diagnosis and Rehabilitation of Autism Spectrum Disorder: A Review. *ArXiv:2007.01285 [Cs, Eess, Stat]*. http://arxiv.org/abs/2007.01285

Khosla, M., Jamison, K., Ngo, G. H., Kuceyeski, A., & Sabuncu, M. R. (2019). Machine learning in resting-state fMRI analysis. *Magnetic Resonance Imaging*, *64*, 101–121. https://doi.org/10.1016/j.mri.2019.05.031

Lau, W. K. W., Leung, M.-K., & Lau, B. W. M. (2019). Resting-state abnormalities in Autism Spectrum Disorders: A meta-analysis. *Scientific Reports*, *9*(1), 3892. https://doi.org/10.1038/s41598-019-40427-7

Li, H., & Fan, Y. (2018). Brain Decoding from Functional MRI Using Long Short-Term Memory Recurrent Neural Networks. In A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, & G. Fichtinger (Eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018* (Vol. 11072, pp. 320–328). Springer International Publishing. https://doi.org/10.1007/978-3-030-00931-1_37

Li, J., Kong, R., Liégeois, R., Orban, C., Tan, Y., Sun, N., Holmes, A. J., Sabuncu, M. R., Ge, T., & Yeo, B. T. T. (2019). Global signal regression strengthens association between resting-state functional connectivity and behavior. *NeuroImage*, *196*, 126–141. https://doi.org/10.1016/j.neuroimage.2019.04.016

Li, X., Zhou, Y., Gao, S., Dvornek, N., Zhang, M., Zhuang, J., Gu, S., Scheinost, D., Staib, L., Ventola, P., & Duncan, J. (2020). BrainGNN: Interpretable Brain Graph Neural Network for fMRI Analysis. *BioRxiv*, 2020.05.16.100057. https://doi.org/10.1101/2020.05.16.100057

Nielsen, J. A., Zielinski, B. A., Fletcher, P. T., Alexander, A. L., Lange, N., Bigler, E. D., Lainhart, J. E., & Anderson, J. S. (2013). Multisite functional connectivity MRI classification of autism: ABIDE results. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00599

Parker, D. B., & Razlighi, Q. R. (2019). The Benefit of Slice Timing Correction in Common fMRI Preprocessing Pipelines. *Frontiers in Neuroscience*, *13*. https://doi.org/10.3389/fnins.2019.00821

Parmar, N., Vaswani, A., Uszkoreit, J., Kaiser, Ł., Shazeer, N., Ku, A., & Tran, D. (2018). Image Transformer. *ArXiv:1802.05751 [Cs]*. http://arxiv.org/abs/1802.05751

Thomas, R. M., Gallo, S., Cerliani, L., Zhutovsky, P., El-Gazzar, A., & van Wingen, G. (2020). Classifying Autism Spectrum Disorder Using the Temporal Statistics of Resting-State Functional MRI Data With 3D Convolutional Neural Networks. *Frontiers in Psychiatry*, *11*. https://doi.org/10.3389/fpsyt.2020.00440

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. *ArXiv:1706.03762 [Cs]*. http://arxiv.org/abs/1706.03762