An Explainable 3D CNN for ASD Diagnosis using structural MRI

Kakyeong Kim1, Yuhyun Cha2, Hyeonjin Kim3

1. **Department of Brain and Cognitive Sciences, College of Natural Sciences, Seoul National University;**
2. **Department of Public Health, School of Public Health, Seoul National University;**
3. **Department of Psychology, College of Social Sciences, Seoul National University;**

Abstract

To date, there is no brain-derived biomarker for autism spectrum disorder (ASD), which leads to delayed or inaccurate diagnosis. Though studies have tried to use machine/deep learning with imaging data for the ASD diagnostic tool, it is hard to be generalized to clinical practice due to small and imbalanced data. Given the problems, we used big data (ABIDE and ABCD) and CT-GAN for the small sample size and imbalanced data between groups, respectively. We show the ASD group in ABIDE is classified by 2D/3D DenseNet. Then we apply transfer-learning to ABCD data after CT-GAN. As a result the better performance is shown from the CT-GAN-generated data set. This indicates the problem of the limited and the imbalanced number of data can be overcome by combining two different machine learning methodologies, CT-GAN with DenseNet.

1. Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder with multiple biological etiologies, characterized by social communication deficits and fixated interests and repetitive behaviors (American Psychiatric Association, 2013). Despite the substantial advances in understanding neurobiology and abnormal brain maturation in ASD, the diagnosis still depends on behavioral tools for assessing social problems such as attentional problems and theory of mind (ToM) (Falkmer et al., 2013), which leads to delayed or inaccurate diagnosis of ASD. For an objective clinical diagnosis and a better understanding of biological mechanisms underlying symptoms, biomarkers should be used. Literature shows that ASD has an atypical trajectory of anatomical and functional brain development. In terms of attention problems, specifically, frontal eye field (FEF), superior colliculus (SC), thalamus, BG have been reported (Esterman et al., 2015; Matsumoto et al., 2018; Xuan et al., 2016). The temporal parietal junction (TPJ) and medial prefrontal cortex (mPFC) have been studied as brain regions for ToM (Filmer et al., 2019). Although volumetric differences have been reported in both gray and white matter, the findings have been inconsistent due to differences in study design, methodology and subject heterogeneity.

Based on the brain abnormalities in ASD, several studies performed a data-driven machine learning approach for classification, prognosis prediction, and treatment response prediction using structural (Kong et al., 2019; Sen et al., 2018) and functional MRI (Dekhil et al., 2018; Eslami et al., 2019; Guo et al., 2017; Heinsfeld et al., 2018; Li et al., 2018) However, it is difficult to generalize the findings to clinical practice due to small sample size and the imbalanced number of clinical data (Eslami et al., 2020), which makes it difficult to optimize the accuracy of the model. To figure out the sampling problem, therefore, firstly, we used the ABCD and ABIDE data which are the largest developmental and international autism brain imaging data, respectively. For the imbalanced data, we utilized CT-GAN to handle the imbalanced number of data between ASD and non-ASD. Furthermore, we attempted to investigate the neuroanatomical features, which contributed to the classification.

2. Related work

2.1. Machine Learning for Neuroimaging-based Diagnosis for ASD

Machine learning has been successfully applied on medical image classification. For neuroimaging-based ASD classification, several studies performed machine learning algorithms. Of the 12 studies using sMRI data reported in the review paper (Moon et al., 2019), 10 studies implemented the analysis with sample size less than 200. Although a higher accuracy (%) up to 98.67 ± 1.7 was reported in ASD classification, it is difficult to generalize the results to other sites due to its small sample size (n=40) (Subbaraju et al., 2015). Indeed, the similar approach using larger sample size shows the poor performance of accuracy (%) 52 ± 7 (n=650) (Demirhan, 2018) and 60 (n=734) (Katuwal et al., 2015). Therefore, it is essential to evaluate the extent to which machine learning can classify large, ecological valid datasets to test the clinical utility of predictive models.

2.2. Working with imbalanced data

In the real world, the major challenge of clinical data is the imbalanced number of data samples. To handle the imbalanced classes, we consider two complementary and alternative methods of deep neural networks: transfer learning and conditional generative adversarial networks (CT-GAN). Transfer learning acquires the knowledge from related tasks to improve generalization in the current interest task (Valverde et al. 2021), so it is beneficial in terms of learning speed and generalization of the neural network when learning data is limited sample size (Park and Ahn 2021). CT-GAN can be utilized to model complex and large-scale dataset by adding conditional variable y to both the generator and discriminator (Alotaibi 2020). In conditional GAN, the model

combines random noise and target conditional variables in the real data x with conditional information y (Alotaibi 2020).

3. Dataset

3.1. ABCD Participants

We obtained the data from the Adolescent Brain Cognitive Development (ABCD) study release 3.0 (http://abcdstudy.org) (Charness, 2018). The ABCD Study is the largest longitudinal study of brain development and child health across the United States. The aim of the ABCD study is to investigate psychological and neurobiological development trajectories for adolescent mental health. 11,875 children aged 9 to 10 years olds were recruited from 21 research sites. After quality control and preprocessing, we used 9,210 participants in this study including 113 ASD participants

3.2. ABIDE Participants

We used Autism Brain Imaging Data Exchange (ABIDE) II data (http:/fcon_1000.projects.nitrc.org/indi/abide) (A. Di Martino et al., 2014; Adriana Di Martino et al., 2017). The ABIDE aims to discover the neural bases of ASD. 1,114 participants (593 non-ASD and 521 ASD) aged 5 to 64 years were recruited from 19 research sites. After quality control and preprocessing, we used 603 participants in this study.

3.3. Image acquisition and preprocessing

We acquired T1-weighted (T1w) 3D structural MRI from the ABCD and ABIDE data (Cameron et al., 2013) preprocessed with fMRIprep (RRID:SCR_016216) version stable (Esteban et al., 2019), a Nipype (RRID:SCR_002502) (K. Gorgolewski et al., 2011; K. J. Gorgolewski et al., 2017) based tool. Each T1w volume was corrected for INU (intensity non-uniformity) using N4BiasFieldCorrection v2.1.0 (Tustison et al., 2010) and skull-stripped using antsBrainExtraction.sh v2.1.0 (using the OASIS template). Spatial normalization to the ICBM PediatricAsymmetrical template (RRID:SCR_008796) (Fonov et al., 2009) was performed through nonlinear registration with the antsRegistration tool of ANTs v2.1.0 (RRID:SCR_004757) (Avants et al., 2008), using brain-extracted versions of both T1w volume and template. Brain tissue segmentation of cerebrospinal fluid (CSF), white matter (WM), and gray matter (GM) were performed on the brain-extracted T1w using fast (Zhang et al., 2001) (FSL v5.0.9, RRID:SCR_002823). The output size of T1wMRI data is 99 x 117 x 95.

4. Methods

4.1. Densely Connected Convolutional Networks (DenseNet)

We performed densely connected convolutional networks (DenseNet) to classify ASD and non-ASD based on T1w 3D structural MRI data. Original 2D DenseNet connects each layer to every other layer in a feed-forward way, which increased direct connections between the low and high layers (Huang et al., 2017). With the direct connection between layers, DenseNet enhances the information flow, strengthens feature propagation, alleviates the vanishing-gradient problem, and reduces the number of parameters (Huang et al., 2017). Our 3D DenseNet model has a similar architecture to the original DenseNet where the number of spatial dimensions of the input image is three instead of two. Preprocessed images (99 x 177 x 95) were resized into 96 x 96 x 96 and fed into 3D DenseNet with 4 dense blocks (*DenseNet121*) (see **Figure 1**). 3D convolution was applied with filter size 7, stride 2, and padding 3, followed by 3D batch normalization and ReLU activation. Then, 3D MaxPool with the kernel size 3, stride 2, and padding 1 was conducted and fed into 4 dense blocks with transition blocks. In each dense block, 1x1x1 3D convolution and 3x3x3 3D convolution filter was used and the number of filters used in dense blocks was (6, 12, 24, 16). In the transition block, 3D batch normalization, ReLU activation, 1x1x1 3D convolution, and 3D AvgPool were applied. After 4 dense blocks with 3 transition blocks, a 3D adaptive AvgPool layer was added to make 1024 1x1x1 dimensions, which is then flattened and fed into a softmax classification layer.

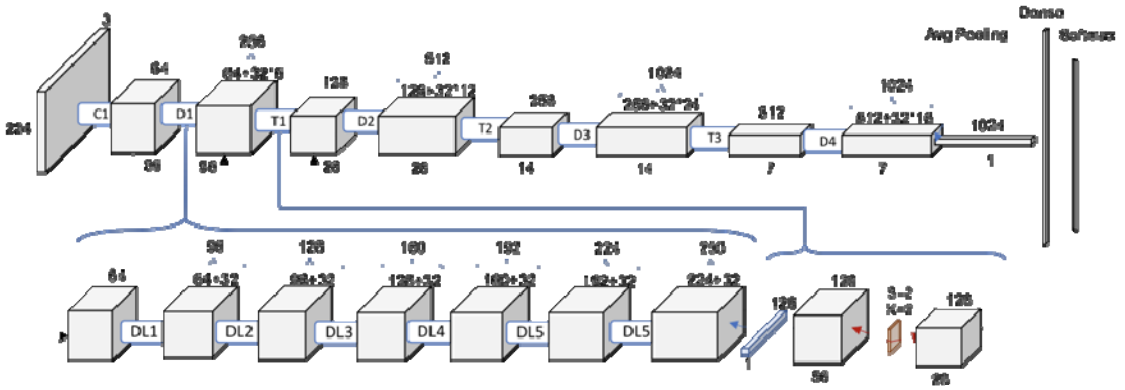


Figure 1: The architecture of DenseNet

4.2. Transfer Learning

Transfer learning is an idea of reusing the pre-trained network on the target dataset, where another large dataset similar to the target dataset was used for the pretraining (Torrey & Shavlik, n.d.). It improves the model performance of the target task by enabling the transfer of knowledge from a similar previous task. This technique is particularly important when it comes to the large brain MRI dataset as it can address the issue of variation of imaging protocols or scanners. Many previous studies utilizing brain images already applied the transfer learning technique (Valverde et al., 2021). We had adopted transfer learning in experiments by pre-training the DenseNet model using ABIDE dataset then applied it to ABCD. We expected improvement of model performance as the ABIDE dataset is more balanced and has a bigger ASD sample size compared to ABCD.

4.3. Conditional Generative Adversarial Networks (CT-GAN)

The generative adversarial networks (GAN) is a deep learning model that has a generative model and discriminative model. A generative model learns the generator's distribution and generates new data, while a discriminative model discriminates between data instances (Goodfellow et al., 2020). With the generated synthetic data, previous studies have successfully performed image generation and object detection with insufficient training data (Douzas & Bacao, 2018; Islam & Zhang, 2020; Isola et al., 2017; Ledig et al., 2017). However, one of the existing limitations of image-to-image translation in this work is the lack of sample size to generate synthetic images. Therefore, we first extracted the value of each brain region from image data as tabular, then applied conditional GAN (CT-GAN) to handle imbalanced data between cases. The CT-GAN is a synthetic data generator for single and complex, multi-table datasets. which can be used to supplement, augment and replace real data (Lei Xu, Maria Skoularidou, Alfredo Cuesta-Infante, Kalyan Veeramachaneni, 2019). With the 6149 non-ASD and 93 ASD samples, we generated 535 ASD data. Furthermore, the CTCAN allows the generation of multi-data types, such as numerical, discrete, and time, so it has been utilized in multimodal medical data.

4.4. Explainable AI (XAI)

Explainable AI (XAI) refers to the learning algorithm where the results can be understood by humans (Sample et al., 2017). Many complex learning algorithms such as Deep Neural Network are regarded as black-box models because humans are not able to understand the underlying mechanism even when they show good model performances. The black-box-ness poses a challenge of trust or justification of conclusions drawn by the models. The challenge exactly puts a big limitation on the medical domain using brain MRI data (e.g. diagnosis) (Holzinger et al., 2017). The complex Deep Neural Networks trained by high-dimensional brain MRI data hamper the application of the results in medical practice since the medical professionals cannot understand how and why a machine decision has been made. To address the problem, we visualize the occlusion sensitivity (Zeiler & Fergus, 2014) for the model's prediction by using a python package, MONAI (Ma et al., 2021). The occlusion map enables us to interpret where the brain area affects certain decisions made by the network. The higher values in the map mean that the occluded region had more impact on the decision process.

5. Experiments and Results

5.1. ABIDE dataset

5.2.1. 2D DenseNet on MRI slices

We conducted 2D DenseNet architecture on the slice of 3T MRI images in ABIDE dataset. We used sagittal images for 2D classification with the size of (193, 193). The algorithm was optimized by Adam optimizer with a learning rate of 1e-3. The final model was selected based on accuracy metrics in the validation set. The results of accuracy metrics for the test dataset are in **Table 1**.

5.1.2. 3D DenseNet on MRI images

Whole 3T MRI images were used to classify ASD patients in ABIDE dataset by the 3D DenseNet algorithm. The original size of images (193, 229, 193) was resized into (96, 96, 96), and the model was optimized with Adam optimizer (learning rate 1e-5). The final model was selected based on accuracy metrics in the validation set, and the results of test data are in **Table 1**.

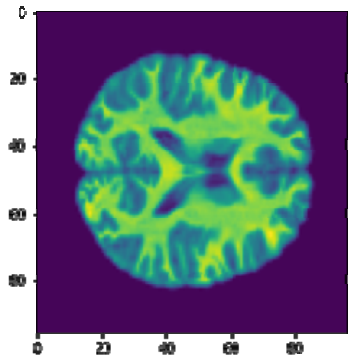
Table 1. Summary of accuracy metrics for the test datasets in 2D DenseNet and 3D DenseNet

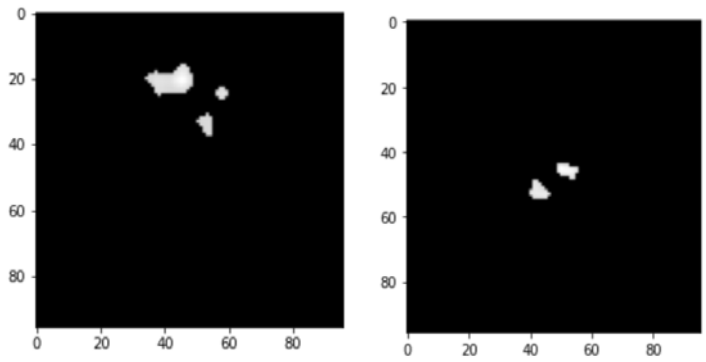
	2D CNN	**3D CNN**
Sensitivity	63%	59%
Specificity	53%	68%
Positive Predictive Value	52%	59%
Negative Predictive Value	64%	68%
Accuracy	**57%**	**64%**

5.1.2.1. Occlusion sensitivity

We computed the occlusion sensitivity for ASD test examples that were successfully classified into ASD groups. Given 3T MRI images, the change of prediction as the mask occludes each part of images was measured. Through occlusion sensitivity map (**Figure 1**), it could be confirmed which part of the brain in ASD participants was particularly used for ASD classification in each sample.

A. An example of axial 3T MRI image





B. Occlusion sensitivity map in ASD examples

Figure 1. Occlusion sensitivity map in ASD

5.2. ABCD dataset

5.2.1. Transfer learning on 3T MRI images

From the model using 3D DenseNet architecture on ABIDE dataset, we also conducted ASD classification in ABCD dataset. Original dimensions of 3T images (99, 117, 95) were resized into (96, 96, 96) and the pre-trained model was optimized with Adam optimizer (learning rate 1e-10). The results of the test dataset were sensitivity 14%, specificity 90%, PPV 23%, NPV 83%, and the total accuracy was 76%.

5.2.2. Data Augmentation using CT-GAN

To handle the unbalanced sample size between non-ASD and ASD, we implemented GAN-based data augmentation with CT-GAN. Based on the training set (non-ASD: 6159, ASD: 93), we generated 535 ASD. We performed baseline classification tasks to compare the results of the proposed frameworks. We used a generalized linear model (GLM), random forest (RF), gradient boosting machine (GBM), and automated machine learning framework before and after CT-GAN in the ABCD dataset. The summary of classification performance using automated ML is shown in **Table 2**. Overall, the original dataset shows better performance than the generated data using CT-GAN. However, synthesized data present higher performance when using a stacked ensemble model (74.77 % of ROC-AUC).

Table 2. Summary of classification performance using automated ML before and after data augmentation with CT-GAN to handle the imbalanced sample set.

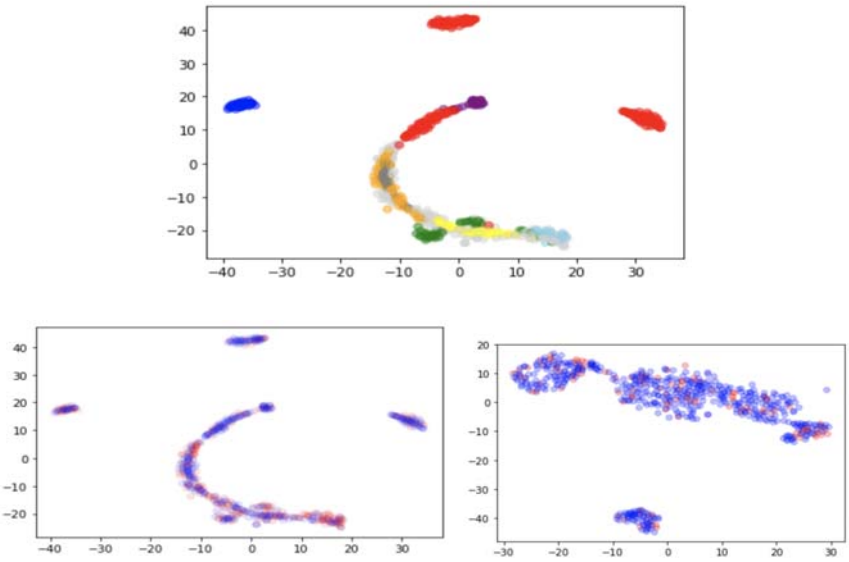
	Before CT-GAN	**After CT-GAN**
Automated machine learning framework	**XGBoost 58.40%**	**Stacked Ensemble 74.77%**
generalized linear model (GLM)	56.76%	49.15%
GLM with lambda search	66.29%	52.06%
random forest (RF) with 50 trees	66.21%	47.24%
RF with 100 trees	64.60%	43.90%
RF with 5-fold cross-validation	60.79	67.24%
gradient boosting machine (GBM) with 50 trees	55.40%	54.00%
GBM with 500 trees	60.67%	47.09%

5.3. Visualization

Through the dimension reduction algorithm (t-SNE), the whole dataset was reduced into 2-dimensional planes, and checked if there is any site-/group-specific difference in each dataset (**Figure 2**). Figure X.A shows

that inter-site difference exists in ABIDE dataset. For ASD classification, ASD group (red) had lower values of t-SNE embeddings compared to non-ASD group (blue)) in ABIDE dataset (**Figure 2**.B; y-axis). However, no difference between ASD and non-ASD groups was observed in the ABCD dataset.

A. Site-specific visualization in ABIDE dataset



B. Group-specific visualization (Left) ABIDE dataset (Right) ABCD dataset
(Red) ASD participants (Blue) Non-ASD participants

Figure 2. t-Stochastic Neighbor Embedding (t-SNE) of ABIDE/ABCD dataset

6. Conclusion

In this paper, we proposed several models for ASD classification from brain imaging data. In particular, we have explored i) 2D slice-level and ii) 3D patch-level DenseNet models in two different datasets, respectively. In the ABCD dataset, we detected a problem of imbalanced sample size between non-ASD and ASD. Therefore, we consider two domain adaptation approaches of transfer learning and CT-GAN. We found that 2D CNN and 3D CNN showed 57% and 64% accuracy, respectively, in the ABIDE dataset. By using transfer learning, the pre-trained 3D CNN in the ABIDE dataset could identify non-ASD children with 90% specificity but ASD children only with 14% sensitivity in the ABCD dataset. However, after synthesizing data produced by CT-GAN (ASD N = 535), the ROC-AUC increased to 74.7% when using a stacked ensemble model in the ABCD dataset.

The main limitation of the research lies in that transfer learning could not increase the model performance. It might indicate that the features of ASD in the brain differ by age group. Since the ABCD dataset only includes a specific age group (9 to 10 years), the learned features from all age groups (5 to 64 years) would not be applicable. Also, the difference in image preprocessing might lead to the result in that the brain data from ABIDE was skull stripped but ABCD was not. Furthermore, although CT-GAN shows 74.7% accuracy, we cannot rule out the possibility that the result of CT-GAN was overfitted due to limited sample size.

In summary, this study reports an acceptable, generalizable performance in predicting ASD with large sample size. To handle imbalanced sample size with transfer learning and CT-GAN, we present a useful framework for future research of studying ASD neuro-biomarker.

References

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub.

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image

registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*(1), 26–41.

Cameron, C., Yassine, B., Carlton, C., Francois, C., Alan, E., András, J., Budhachandra, K., John, L., Qingyang, L., Michael, M., Chaogan, Y., & Pierre, B. (2013). The Neuro Bureau Preprocessing Initiative: open sharing of preprocessed neuroimaging data and derivatives. In *Frontiers in Neuroinformatics* (Vol. 7). https://doi.org/10.3389/conf.fninf.2013.09.00041

Charness, M. E. (2018). The adolescent brain cognitive development study external advisory board. In *Developmental Cognitive Neuroscience* (Vol. 32, pp. 155–160). https://doi.org/10.1016/j.dcn.2017.12.007

Dekhil, O., Hajjdiab, H., Shalaby, A., Ali, M. T., Ayinde, B., Switala, A., Elshamekh, A., Ghazal, M., Keynton, R., Barnes, G., & El-Baz, A. (2018). Using resting state functional MRI to build a personalized autism diagnosis system. *PloS One*, *13*(10), e0206351.

Demirhan, A. (2018). The effect of feature selection on multivariate pattern analysis of structural brain MR images. *Physica Medica: PM: An International Journal Devoted to the Applications of Physics to Medicine and Biology: Official Journal of the Italian Association of Biomedical Physics* , *47*, 103–111.

Di Martino, A., O'Connor, D., Chen, B., Alaerts, K., Anderson, J. S., Assaf, M., Balsters, J. H., Baxter, L., Beggiato, A., Bernaerts, S., Blanken, L. M. E., Bookheimer, S. Y., Braden, B. B., Byrge, L., Castellanos, F. X., Dapretto, M., Delorme, R., Fair, D. A., Fishman, I., … Milham, M. P. (2017). Enhancing studies of the connectome in autism using the autism brain imaging data exchange II. *Scientific Data*, *4*, 170010.

Di Martino, A., Yan, C.-G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., Anderson, J. S., Assaf, M., Bookheimer, S. Y., Dapretto, M., Deen, B., Delmonte, S., Dinstein, I., Ertl-Wagner, B., Fair, D. A., Gallagher, L., Kennedy, D. P., Keown, C. L., Keysers, C., … Milham, M. P. (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular Psychiatry*, *19*(6), 659–667.

Douzas, G., & Bacao, F. (2018). Effective data generation for imbalanced learning using conditional generative adversarial networks. In *Expert Systems with Applications* (Vol. 91, pp. 464–471). https://doi.org/10.1016/j.eswa.2017.09.030

Eslami, T., Almuqhim, F., Raiker, J. S., & Saeed, F. (2020). Machine Learning Methods for

Diagnosing Autism Spectrum Disorder and Attention- Deficit/Hyperactivity Disorder Using Functional and Structural MRI: A Survey. *Frontiers in Neuroinformatics*, *14*, 575999.

Eslami, T., Mirjalili, V., Fong, A., Laird, A. R., & Saeed, F. (2019). ASD-DiagNet: A Hybrid Learning Approach for Detection of Autism Spectrum Disorder Using fMRI Data. *Frontiers in Neuroinformatics*, *13*, 70.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116.

Esterman, M., Liu, G., Okabe, H., Reagan, A., Thai, M., & DeGutis, J. (2015). Frontal eye field involvement in sustaining visual attention: evidence from transcranial magnetic stimulation. *NeuroImage*, *111*, 542–548.

Falkmer, T., Anderson, K., Falkmer, M., & Horlin, C. (2013). Diagnostic procedures in autism spectrum disorders: a systematic literature review. In *European Child & Adolescent Psychiatry* (Vol. 22, Issue 6, pp. 329–340). https://doi.org/10.1007/s00787-013-0375-0

Filmer, H. L., Fox, A., & Dux, P. E. (2019). Causal evidence of right temporal parietal junction involvement in implicit Theory of Mind processing. *NeuroImage*, *196*, 329–336.

Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R., & Collins, D. L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, *47*, S102.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. In *Communications of the ACM* (Vol. 63, Issue 11, pp. 139–144). https://doi.org/10.1145/3422622

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in Neuroinformatics*, *5*, 13.

Gorgolewski, K. J., Esteban, O., Ellis, D. G., Notter, M. P., Ziegler, E., Johnson, H., Hamalainen, C., Yvernault, B., Burns, C., Manhães-Savio, A., Jarecka, D., Markiewicz, C. J., Salo, T., Clark, D., Waskom, M., Wong, J., Modat, M., Dewey, B. E., Clark, M. G., … Ghosh, S. (2017). *Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python. 0.13.1*. https://doi.org/10.5281/zenodo.581704

Guo, X., Dominick, K. C., Minai, A. A., Li, H., Erickson, C. A., & Lu, L. J. (2017). Diagnosing Autism Spectrum Disorder from Brain Resting-State Functional Connectivity Patterns Using a Deep Neural Network with a Novel Feature Selection Method. In *Frontiers in Neuroscience* (Vol. 11). https://doi.org/10.3389/fnins.2017.00460

Heinsfeld, A. S., Franco, A. R., Craddock, R. C., Buchweitz, A., & Meneguzzi, F. (2018). Identification of autism spectrum disorder using deep learning and the ABIDE dataset. *NeuroImage. Clinical*, *17*, 16–23.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4700–4708.

Islam, J., & Zhang, Y. (2020). GAN-based synthetic brain PET image generation. *Brain Informatics*, *7*(1), 3.

Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Image-to-Image Translation with Conditional Adversarial Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr.2017.632

Katuwal, G. J., Cahill, N. D., Baum, S. A., & Michael, A. M. (2015). The predictive power of structural MRI in Autism diagnosis. *Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, *2015*, 4270–4273.

Kong, Y., Gao, J., Xu, Y., Pan, Y., Wang, J., & Liu, J. (2019). Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier. In *Neurocomputing* (Vol. 324, pp. 63–68). https://doi.org/10.1016/j.neucom.2018.04.080

Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2017). Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr.2017.19

Lei Xu, Maria Skoularidou, Alfredo Cuesta-Infante, Kalyan Veeramachaneni. (2019). *Synthesizing Tabular Data Using Conditional GAN*. https://arxiv.org/abs/1907.00503

Li, G., Liu, M., Sun, Q., Shen, D., & Wang, L. (2018). Early Diagnosis of Autism Disease by Multi-channel CNNs. *Machine Learning in Medical Imaging. MLMI (Workshop)*, *11046*, 303–309.

Matsumoto, M., Inoue, K.-I., & Takada, M. (2018). Causal Role of Neural Signals Transmitted From the Frontal Eye Field to the Superior Colliculus in Saccade Generation. *Frontiers in Neural Circuits*, *12*, 69.

Moon, S. J., Hwang, J., Kana, R., Torous, J., & Kim, J. W. (2019). Accuracy of Machine Learning Algorithms for the Diagnosis of Autism Spectrum Disorder: Systematic Review and Meta-Analysis of Brain Magnetic Resonance Imaging Studies. In *JMIR Mental Health* (Vol. 6, Issue 12, p. e14108). https://doi.org/10.2196/14108

Sen, B., Borle, N. C., Greiner, R., & Brown, M. R. G. (2018). A general prediction model for the detection of ADHD and Autism using structural and functional MRI. In *PLOS ONE* (Vol. 13, Issue 4, p. e0194856). https://doi.org/10.1371/journal.pone.0194856

Subbaraju, V., Sundaram, S., Narasimhan, S., & Suresh, M. B. (2015). Accurate detection of autism spectrum disorder from structural MRI using extended metacognitive radial basis function network. In *Expert Systems with Applications* (Vol. 42, Issue 22, pp. 8775–8790). https://doi.org/10.1016/j.eswa.2015.07.031

Torrey, L., & Shavlik, J. (n.d.). Transfer Learning. In *Handbook of Research on Machine Learning Applications and Trends* (pp. 242–264). https://doi.org/10.4018/978-1-60566-766-9.ch011

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: improved N3 bias correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–1320.

Valverde, J. M., Imani, V., Abdollahzadeh, A., De Feo, R., Prakash, M., Ciszek, R., & Tohka, J. (2021). Transfer Learning in Magnetic Resonance Brain Imaging: A Systematic Review. In *Journal of Imaging* (Vol. 7, Issue 4, p. 66). https://doi.org/10.3390/jimaging7040066

Xuan, B., Mackie, M.-A., Spagna, A., Wu, T., Tian, Y., Hof, P. R., & Fan, J. (2016). The activation of interactive attentional networks. *NeuroImage*, *129*, 308–319.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. In *Computer Vision – ECCV 2014* (pp. 818–833). https://doi.org/10.1007/978-3-319-10590-1_53

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45–57.

American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). American Psychiatric Pub.

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Medical Image Analysis, 12(1), 26–41.

Charness, M. E. (2018). The adolescent brain cognitive development study external advisory board. In Developmental Cognitive Neuroscience (Vol. 32, pp. 155–160). https://doi.org/10.1016/j.dcn.2017.12.007

Dekhil, O., Hajjdiab, H., Shalaby, A., Ali, M. T., Ayinde, B., Switala, A., Elshamekh, A., Ghazal, M., Keynton, R., Barnes, G., & El-Baz, A. (2018). Using resting state functional MRI to build a personalized autism diagnosis system. PloS One, 13(10), e0206351.

Di Martino, A., O'Connor, D., Chen, B., Alaerts, K., Anderson, J. S., Assaf, M., Balsters, J. H., Baxter, L., Beggiato, A., Bernaerts, S., Blanken, L. M. E., Bookheimer, S. Y., Braden, B. B., Byrge, L., Castellanos, F. X., Dapretto, M., Delorme, R., Fair, D. A., Fishman, I., … Milham, M. P. (2017). Enhancing studies of the connectome in autism using the autism brain imaging data exchange II. Scientific Data, 4, 170010.

Di Martino, A., Yan, C.-G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., Anderson, J. S., Assaf, M., Bookheimer, S. Y., Dapretto, M., Deen, B., Delmonte, S., Dinstein, I., Ertl-Wagner, B., Fair, D. A., Gallagher, L., Kennedy, D. P., Keown, C. L., Keysers, C., … Milham, M. P. (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular Psychiatry, 19(6), 659–667.

Eslami, T., Almuqhim, F., Raiker, J. S., & Saeed, F. (2020). Machine Learning Methods for Diagnosing Autism Spectrum Disorder and Attention- Deficit/Hyperactivity Disorder Using Functional and Structural MRI: A Survey. Frontiers in Neuroinformatics, 14, 575999.

Eslami, T., Mirjalili, V., Fong, A., Laird, A. R., & Saeed, F. (2019). ASD-DiagNet: A Hybrid Learning Approach for Detection of Autism Spectrum Disorder Using fMRI Data. Frontiers in Neuroinformatics, 13, 70.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R.

A., & Gorgolewski, K. J. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. Nature Methods, 16(1), 111–116.

Esterman, M., Liu, G., Okabe, H., Reagan, A., Thai, M., & DeGutis, J. (2015). Frontal eye field involvement in sustaining visual attention: evidence from transcranial magnetic stimulation. NeuroImage, 111, 542–548.

Falkmer, T., Anderson, K., Falkmer, M., & Horlin, C. (2013). Diagnostic procedures in autism spectrum disorders: a systematic literature review. In European Child & Adolescent Psychiatry (Vol. 22, Issue 6, pp. 329–340). https://doi.org/10.1007/s00787-013-0375-0

Filmer, H. L., Fox, A., & Dux, P. E. (2019). Causal evidence of right temporal parietal junction involvement in implicit Theory of Mind processing. NeuroImage, 196, 329–336.

Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R., & Collins, D. L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage, 47, S102.

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. Frontiers in Neuroinformatics, 5, 13.

Gorgolewski, K. J., Esteban, O., Ellis, D. G., Notter, M. P., Ziegler, E., Johnson, H., Hamalainen, C., Yvernault, B., Burns, C., Manhães-Savio, A., Jarecka, D., Markiewicz, C. J., Salo, T., Clark, D., Waskom, M., Wong, J., Modat, M., Dewey, B. E., Clark, M. G., … Ghosh, S. (2017). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python. 0.13.1. https://doi.org/10.5281/zenodo.581704

Guo, X., Dominick, K. C., Minai, A. A., Li, H., Erickson, C. A., & Lu, L. J. (2017). Diagnosing Autism Spectrum Disorder from Brain Resting-State Functional Connectivity Patterns Using a Deep Neural Network with a Novel Feature Selection Method. In Frontiers in Neuroscience (Vol. 11). https://doi.org/10.3389/fnins.2017.00460

Heinsfeld, A. S., Franco, A. R., Craddock, R. C., Buchweitz, A., & Meneguzzi, F. (2018). Identification of autism spectrum disorder using deep learning and the ABIDE dataset. NeuroImage. Clinical, 17, 16–23.

Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2017). What do we need to build explainable AI systems for the medical domain?. arXiv preprint arXiv:1712.09923.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional

networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4700–4708.

Kong, Y., Gao, J., Xu, Y., Pan, Y., Wang, J., & Liu, J. (2019). Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier. In Neurocomputing (Vol. 324, pp. 63–68). https://doi.org/10.1016/j.neucom.2018.04.080

Li, G., Liu, M., Sun, Q., Shen, D., & Wang, L. (2018). Early Diagnosis of Autism Disease by Multi-channel CNNs. Machine Learning in Medical Imaging. MLMI (Workshop), 11046, 303–309.

Matsumoto, M., Inoue, K.-I., & Takada, M. (2018). Causal Role of Neural Signals Transmitted From the Frontal Eye Field to the Superior Colliculus in Saccade Generation. Frontiers in Neural Circuits, 12, 69.

Nic Ma, Wenqi Li, Richard Brown, Yiheng Wang, Behrooz, Benjamin Gorman, … cgrain. (2021, June 1). Project-MONAI/MONAI: 0.5.3 (Version 0.5.3). Zenodo. http://doi.org/10.5281/zenodo.4891800

Sample, I. (2017). Computer says no: why making ais fair, accountable and transparent is crucial. The Guardian, 5, 1-15.

Sen, B., Borle, N. C., Greiner, R., & Brown, M. R. G. (2018). A general prediction model for the detection of ADHD and Autism using structural and functional MRI. In PLOS ONE (Vol. 13, Issue 4, p. e0194856). https://doi.org/10.1371/journal.pone.0194856

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: improved N3 bias correction. IEEE Transactions on Medical Imaging, 29(6), 1310–1320.

Xuan, B., Mackie, M.-A., Spagna, A., Wu, T., Tian, Y., Hof, P. R., & Fan, J. (2016). The activation of interactive attentional networks. NeuroImage, 129, 308–319.

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Transactions on Medical Imaging, 20(1), 45–57.