Copy Number Variation Informs fMRI-based Prediction of Autism Spectrum Disorder

Nicha C. Dvornek[1,2], Catherine Sullivan[3], James S. Duncan[1,2], and Abha R. Gupta[3]

[1] Department of Radiology & Biomedical Imaging, Yale School of Medicine, New Haven, CT 06510, USA
[2] Department of Biomedical Engineering, Yale University, New Haven, CT 06511, USA
[3] Department of Pediatrics, Yale School of Medicine, New Haven, CT 06510, USA
{nicha.dvornek, catherine.sullivan, james.duncan, abha.gupta}@yale.edu

Abstract. The multifactorial etiology of autism spectrum disorder (ASD) suggests that its study would benefit greatly from multimodal approaches that combine data from widely varying platforms, e.g., neuroimaging, genetics, and clinical characterization. Prior neuroimaging-genetic analyses often apply naive feature concatenation approaches in data-driven work or use the findings from one modality to guide posthoc analysis of another, missing the opportunity to analyze the paired multimodal data in a truly unified approach. In this paper, we develop a more integrative model for combining genetic, demographic, and neuroimaging data. Inspired by the influence of genotype on phenotype, we propose using an attention-based approach where the genetic data guides attention to neuroimaging features of importance for model prediction. The genetic data is derived from copy number variation parameters, while the neuroimaging data is from functional magnetic resonance imaging. We evaluate the proposed approach on ASD classification and severity prediction tasks, using a sex-balanced dataset of 228 ASD and typically developing subjects in a 10-fold cross-validation framework. We demonstrate that our attention-based model combining genetic information, demographic data, and functional magnetic resonance imaging results in superior prediction performance compared to other multimodal approaches.

Keywords: fMRI, Genetics, Multimodal analysis, Autism spectrum disorder

1 Introduction

Autism spectrum disorder (ASD) is characterized by impaired communication and social skills, and restricted, repetitive, and stereotyped behaviors that result in significant disability [2]. ASD refers to a spectrum of disorders due to its heterogeneity, with multiple etiologies, sub-types, and developmental trajectories [20], resulting in diverse clinical presentation of symptoms and severity. Two major factors contributing to the heterogeneity of ASD include *genetic variability*

and *sex* [20]. Research aimed at uncovering the pathophysiology of ASD and its heterogeneous presentations is critical to reduce disparities in early diagnosis and develop personalized targeted treatments.

A popular data-driven approach for discovering biomarkers for ASD is to first build a classification model that can distinguish ASD vs. typically developing (TD) individuals. Prior work generally focuses on unimodal data, e.g., functional magnetic resonance imaging (fMRI), structural MRI, genetics, or behavioral or developmental scores alone [13]. However, given the multifactorial etiology of ASD [19], a unified multimodal approach should improve model classification performance. Furthermore, convergence between different modality datasets on where in the brain ASD may arise and which brain regions correlate with different clinical measures would provide greater confidence in the results.

Prior multimodal methods combining genetic and neuroimaging data often naively concatenate the multimodal features and use them as inputs in a machine learning algorithm [4,26,11]. We aim to integrate the multimodal data in a more informative model design. Furthermore, such concatenation approaches may suffer from differing scales of the multimodal data, which will artificially give one modality greater importance. Another major direction is to use the findings from one modality to guide posthoc analysis of another [14,3,25], missing the opportunity to analyze the paired multimodal data in a truly unified approach. While phenotypic data has been combined with neuroimaging data in multiple ASD studies using thoughtful model designs [21,10,8], such integration between genetic and neuroimaging data has not been explored.

Here, we propose to improve characterization of the neurobiology of ASD by developing an integrated neuroimaging-genetic deep learning model to accurately classify ASD vs. TD individuals and predict ASD severity. As individual genetic differences will influence neuroimaging phenotypes, we propose using an attention-based approach where genetic variables inform what neuroimaging features should be attended to for model prediction. We assess the performance of the proposed approach on ASD classification and severity prediction tasks, using a sex-balanced dataset of 228 ASD and TD subjects in a 10-fold cross-validation framework. We demonstrate superior performance to other methods of combining multimodal data.

2   Methods

2.1   Dataset and preprocessing

The dataset includes a sex-balanced cohort of 228 youth (age range: $8.0 - 17.9$ years), 114 with ASD (59 female, 55 male) and 114 TD controls (58 female, 56 male), available from NIH NDA collection 2021. Data types utilized include clinical measures, genome-wide genotyping, and neuroimaging.

Clinical measures utilized include age at time of scan, sex, ASD diagnosis, and Social Responsiveness Scale-2 (SRS, range: $1 - 162$) [7]. SRS is a measure of severity of social impairment in ASD, but was assessed on both ASD and TD subjects.
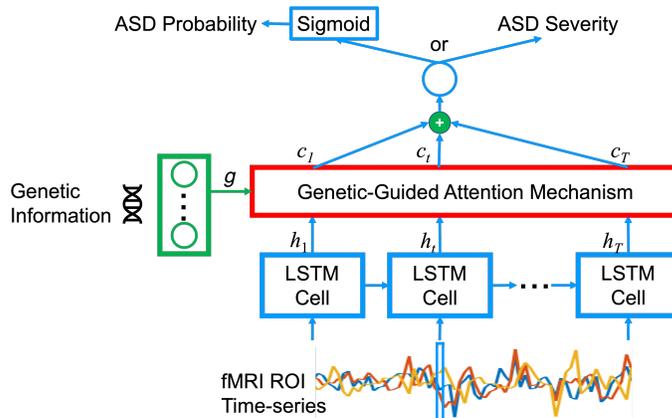
Fig. 1. The proposed integrated neuroimaging- genetic RNN model. The genetic information will be used to focus attention to different fMRI features for the prediction task.

Genome-wide genotyping data was generated using the HumanOmni 2.5M BeadChip (Illumina). This data was processed and analyzed for rare ($>50\%$ of copy number variants (CNV) at $<1\%$ frequency in the Database of Genomic Variants) genic CNVs [14]. CNV parameters of number of CNVs (range: $0-10$) and summed total length of all CNVs (range: $0-4050194$ bp) were used as model inputs.

Structural and fMRI data were acquired under several different tasks. We utilized the Biopoint task-fMRI acquisition, in which subjects viewed coherent and scrambled biological point-light animations in alternating blocks (24s per block; 154 volumes; TR = 2000 ms; TE = 30 ms; flip angle = 90°; FOV = 192 mm; image matrix = 64 mm^2; voxel size = 3 mm × 3 mm × 4 mm; 34 slices). Prior Biopoint studies have identified dysfunction in biological motion processing as reflecting key neural signatures of ASD and as a neuroendophenotype of genetic risk in unaffected siblings [17,5]. fMRI data was preprocessed in FSL [16], including MCFLIRT motion correction, interleaved slice timing correction, BET brain extraction, spatial smoothing (FWHM 5mm), high-pass temporal filtering, and registration to Montreal Neurological Institute space. We parcellated the brain into 116 regions of interest (ROIs) with the AAL atlas [23]. Standardized mean time-series from each ROI were used as model inputs.

2.2   Network architecture

The structure of the proposed genetic-neuroimaging model is shown in Fig. 1. The fMRI time-series from predefined ROIs is first input to a long short-term memory (LSTM) layer [12] to encode fMRI information [9]. Then, inspired by the influence of genotype on phenotype, we propose using a generalized attention mechanism [24] to steer focus to different fMRI features according to the genetic

information. We utilize the genetic data of CNV number and total CNV size derived from genome-wide genotyping, since larger CNV size has been associated with deleteriousness in previous ASD genetics studies [15,22].

The generalized attention mechanism can be defined as a mapping between a query and key-value pair to the "context". Here, we define the query by the genetic and demographic data $g \in R^G$, and the key and value are defined by the encoded fMRI data, which are the outputs of the LSTM $h_t \in R^L$. Applying scaled dot product attention [24], the context is computed as

$$c_t = att(g, h_t) = softmax \left[\frac{(W_q g)^T (W_k h_t)}{\sqrt{M}} \right] W_v h_t, \qquad (1)$$

where $softmax(a_t)$ normalizes a_t such that $\sum_{t=1}^T a_t = 1$, W_q encodes the genetic and demographic information g into the query, W_k and W_v encode fMRI-based h_t into the key and value, respectively, and M denotes the dimension of the encoded space. Multiple attention mappings with different encodings W could be learned to diversify ways in which genetic information modulates the fMRI information. We then summarize the context across the T timesteps $\sum_{t=1}^T c_t$ and apply a fully connected layer to predict the output from the summary M context features. The ASD classification model ends with a sigmoid activation function to produce the probability of the ASD class, while the ASD severity regression model ends with the output of the fully connected layer.

2.3   Experimental Settings

To assess the effectiveness of our multimodal attention-based approach (*Att*), we compared to 4 other methods:
1) *Base*: A basic LSTM model using only the fMRI time-series data [9].
2) *Concat*: Concatenation of the genetic and demographic data with the fMRI time-series data. The genetic and demographic values are repeated across time and concatenated as additional features to the fMRI data. The data are then input to the basic LSTM model.
3) *Fusion*: Fusion of the genetic and demographic information with the prediction from the LSTM processed fMRI data [8]. The genetic and demographic information is combined with the LSTM block using a fully connected layer, followed by a layer with a single node for the task prediction.
4) *LSTMinit*: The baseline LSTM model with initialization of hidden states conditioned on the genetic and demographic variables [10]. The LSTM initial state vectors are learned using a single fully connected layer with size L.

We also perform an ablation study to assess the utility of including both genetic and demographic information to guide the attention to fMRI features in the proposed full model. We thus trained reduced models that included only demographic (*Att-demo*) or genetic (*Att-gene*) information alone to evaluate their value.

All models were implemented in Python using Keras [6] and Tensorflow [1] libraries. The feature dimension for the LSTM output was $L = 16$, and the

feature dimension for the attention embedding was set to $M = 8$. fMRI time-series for each ROI were standardized to have mean 0 and standard deviation 1, and genetic and demographic inputs were normalized to the range [-1,1]. SRS raw scores were normalized to the range [0,1]. We applied randomly shifted windowed samples of the fMRI time-series data as data augmentation [9], sampling 10 random windows of size $T = 48$ (representing the time for 2 blocks each of scrambled and biological motion) per subject every epoch. We used the Adam optimizer [18] ($lr = 0.001$) to train the models for up to 50 epochs. Classification models were trained to optimize binary cross-entropy loss and severity regression models were trained to optimize mean squared error loss.

To evaluate model performance, we performed 10-fold cross-validation of subjects, using stratified sampling of ASD status, with 80% subjects for training, 10% for validation, and 10% for testing in each partition. Validation loss was used to determine the stopping epoch for model selection. For classification models, we computed classification accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC) for each fold. We also computed the overall AUC based on aggregating the test predictions from all folds. For regression models, we computed the mean squared error (MSE), maximum of the mean squared error, and Pearson correlation between true and predicted test outputs for each fold. Because Pearson correlation can vary largely when there are small perturbations in a small test dataset, we also computed the overall Pearson correlation based on aggregating the test predictions from all folds. Significant differences between models were assessed using paired two-tailed t-tests to compare matched cross-validation folds ($\alpha = 0.05$).

3   Results

3.1   ASD Classification

Classification performance ASD classification model results are summarized in Table 1, and receiver operating characteristic curves are plotted in Fig. 2. Traditional concatenation of multimodal inputs performed worse than the fMRI-only base model. Fusing multimodal feature in later layers produced the highest sensitivity, but was not significantly different from the base or our attention-based model results. Initializing the LSTM with genetic and demographic information performed similarly to the fMRI-only base model. Our attention-based approach using the genetic information to inform the attending of important fMRI features resulted in the highest classification accuracy, specificity, and AUC. Furthermore, our attention-based model is the only multimodal approach to perform significantly better than the fMRI-only base model based on the accuracy metric. Moreover, our approach significantly outperformed the other multimodal models as measured by accuracy, specificity, and AUC. As seen in Fig. 2, the receiver operating characteristic curve for our attention model lies above the other models for much of the plot.

When training the attention-based model with only demographic or genetic data for guidance, the models still produced generally better results than the

Table 1. ASD classification performance (mean ± standard deviation). Best results marked in bold.

Model	Accuracy	Sensitivity	Specificity	AUC	Overall AUC
Base	$0.59 \pm 0.13^*$	0.67 ± 0.19	0.52 ± 0.22	0.63 ± 0.14	0.62
Concat	$0.54 \pm 0.07^*$	0.64 ± 0.14	$0.44 \pm 0.15^*$	$0.56 \pm 0.12^{*\dagger}$	0.56
Fusion	$0.60 \pm 0.12^*$	$\mathbf{0.73 \pm 0.11}$	$0.46 \pm 0.20^*$	0.63 ± 0.13	0.62
LSTMinit	0.62 ± 0.11	0.69 ± 0.12	$0.54 \pm 0.12^*$	0.64 ± 0.14	0.63
Att-demo	0.68 ± 0.06	0.67 ± 0.13	0.68 ± 0.11	$\mathbf{0.70 \pm 0.08}$	0.69
Att-gene	0.67 ± 0.06	0.68 ± 0.11	$0.65 \pm 0.13^*$	0.69 ± 0.08	0.69
Att (ours)	$\mathbf{0.69 \pm 0.06^{\dagger}}$	0.68 ± 0.13	$\mathbf{0.69 \pm 0.11}$	$\mathbf{0.70 \pm 0.08}$	**0.70**

* Significant difference compared to our proposed approach Att ($p < 0.05$, paired two-tailed t-test)
† Significant difference compared to fMRI-only Base model ($p < 0.05$, paired two-tailed t-test)

non-attention models, but with a small performance drop compared to the full model utilizing both demographic and genetic information. Using genetic information alone resulted in significantly worse specificity compared to the full model. Again, only the full model utilizing both genetic and demographic data performed significantly better than the unimodal fMRI base model (as measured by accuracy), suggesting the importance of including all modalities in computing the attention scores.
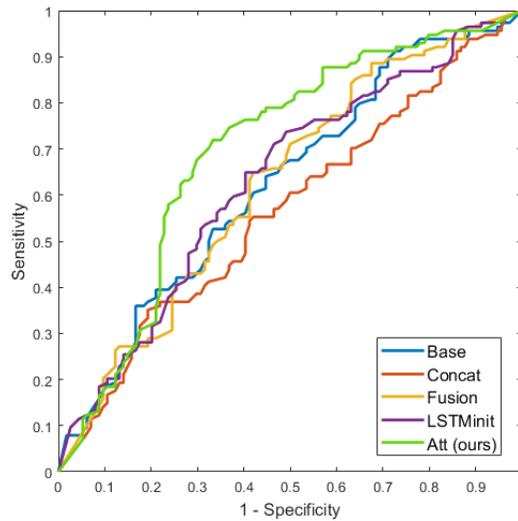
Fig. 2. Receiver operating characteristic curves from ASD classification models.
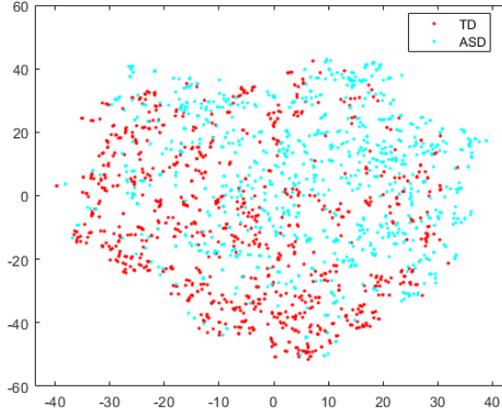
Fig. 3. tSNE visualization of the summary context vector for each sample.
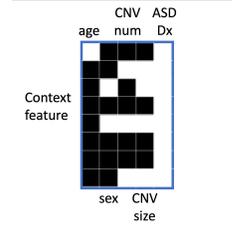
Fig. 4. Context features are associated with different genetic and demographic features.

Model analysis We investigated our attention-based model's representation of multimodal information via the summary context vector. First, we performed a tSNE visualization of the context for one trained model (Fig. 3), where TD subjects are plotted in red and ASD subjects are plotted in blue. We see that there is clustering of the TD and ASD samples.

We then investigate the impact of each genetic and demographic variable on each context feature. We compute the Pearson correlation between a genetic or demographic variable $g(i)$ and the ASD diagnosis and each feature of the context representation $c(j)$ across all samples in a representative fold. We visualize the input and output features with significant correlations ($p < 0.05$) for each context feature by white boxes in Fig. 4. As expected, every context feature is correlated with the ASD diagnosis. Furthermore, we see that different context features are guided by different input features. For example, the context feature corresponding to row 1 has only 1 feature, i.e., age, associated with it, while the second context feature is associated with both genetic variables. Our attention-based approach allows for different modalities of inputs to be integrated in a dynamic manner to produce more informative, individualized representations of the input data, which lends to the improved classification performance.

3.2 ASD Severity Regression

ASD severity regression results are summarized in Table 2. The fMRI-only base model was able to achieve a significant overall correlation. Concatenating genetic and demographic features with fMRI nominally improved the prediction performance. Fusing the genetic and demographic features with fMRI features significantly reduced the MSE compared to the fMRI-only base model and produced the lowest maximum errors; however, the correlation between predicted

Table 2. ASD severity prediction performance (normalized SRS mean $\pm$ standard deviation). Best results marked in bold.

Model	MSE	Maximum	Correlation	Overall Correlation
Base	$0.084 \pm 0.014^*$	0.30 ± 0.09	0.16 ± 0.19	$0.15^\star$
Concat	0.080 ± 0.016	0.29 ± 0.06	0.17 ± 0.21	$0.16^\star$
Fusion	$0.075 \pm 0.016^\dagger$	$\mathbf{0.26 \pm 0.08}$	0.08 ± 0.12	0.08
LSTMinit	0.087 ± 0.011	0.28 ± 0.08	0.13 ± 0.28	0.12
Att-demo	$\mathbf{0.074 \pm 0.019}^\dagger$	0.29 ± 0.09	$\mathbf{0.19 \pm 0.11}$	$\mathbf{0.18}^{\star\star}$
Att-gene	$\mathbf{0.074 \pm 0.019}^\dagger$	0.29 ± 0.08	$\mathbf{0.19 \pm 0.10}$	$0.16^\star$
Att (ours)	$\mathbf{0.074 \pm 0.018}^\dagger$	0.29 ± 0.09	$\mathbf{0.19 \pm 0.12}$	$\mathbf{0.18}^{\star\star}$

* Significant difference compared to our proposed approach Att ($p < 0.05$, paired two-tailed t-test)

† Significant difference compared to fMRI-only Base model ($p < 0.05$, paired two-tailed t-test)

* / ** Significant difference from 0 ($p < 0.05$ / $p < 0.01$, two-tailed t-test)

and true scores was greatly reduced and no longer significant. The model initializing the LSTM with genetic and demographic variables performed slightly worse than the base fMRI model. Our attention-based approach resulted in the lowest MSE and highest correlation.

As seen in the ablation study results, training the attention module with demographic or genetic data alone produces similar results to training with both modes of data together in our full model. All attention-based models performed significantly better than the unimodal fMRI model as measured by MSE. Using genetic data alone produced a slightly lower overall correlation. Although the performance levels are similar, including both genetic and demographic information may result in better understand of the interplay between these different types of individual features in the context of understanding ASD.

4   Conclusions

In this work, we proposed an attention-based approach to integrating genetic and demographic information with fMRI data. The genetic and demographic data are used to guide attention to important fMRI encoded features. In a 10-fold cross-validation framework with a dataset of 228 ASD and TD subjects, We demonstrated improved performance in an ASD classification and ASD severity prediction task compared to other standard approaches of combining multimodal data. Future work will explore other genetic variables to use as input (e.g., number of genes contained in CNVs) and analyze the neuroimaging biomarkers that are used by the attention-based model for ASD classification and severity prediction.

References

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.: TensorFlow: Large-scale machine learning on heterogeneous systems (2015), https://www.tensorflow.org/, software available from tensorflow.org
2. American Psychiatric Association, Washington, DC: Diagnostic and Statistical Manual of Mental Disorders, 5th edn. (2013)
3. Antonucci, L.A., Di Carlo, P., Passiatore, R., Papalino, M., Monda, A., Amoroso, N., Tangaro, S., Taurisano, P., Rampino, A., Sambataro, F., et al.: Thalamic connectivity measured with fmri is associated with a polygenic index predicting thalamo-prefrontal gene co-expression. Brain Structure and Function **224**, 1331–1344 (2019)
4. Bi, X.a., Hu, X., Xie, Y., Wu, H.: A novel cernne approach for predicting parkinson's disease-associated genes and brain regions based on multimodal imaging genetics data. Medical Image Analysis **67**, 101830 (2021)
5. Björnsdotter, M., Wang, N., Pelphrey, K., Kaiser, M.D.: Evaluation of quantified social perception circuit activity as a neurobiological marker of autism spectrum disorder. JAMA psychiatry **73**(6), 614–621 (2016)
6. Chollet, F., et al.: Keras (2015), software available on Github
7. Constantino, J.N., Gruber, C.P.: Social responsiveness scale: SRS-2. Western psychological services Torrance, CA (2012)
8. Dvornek, N.C., Ventola, P., Duncan, J.S.: Combining phenotypic and resting-state fmri data for autism classification with recurrent neural networks. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). pp. 725–728. IEEE (2018)
9. Dvornek, N.C., Ventola, P., Pelphrey, K.A., Duncan, J.S.: Identifying autism from resting-state fmri using long short-term memory networks. In: Machine Learning in Medical Imaging: 8th International Workshop, MLMI 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 10, 2017, Proceedings 8. pp. 362–370. Springer (2017)
10. Dvornek, N.C., Yang, D., Ventola, P., Duncan, J.S.: Learning generalizable recurrent neural networks from small task-fmri datasets. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 329–337. Springer (2018)
11. Heinrich, A., Müller, K.U., Banaschewski, T., Barker, G.J., Bokde, A.L., Bromberg, U., Büchel, C., Conrod, P., Fauth-Bühler, M., Papadopoulos, D., et al.: Prediction of alcohol drinking in adolescents: Personality-traits, behavior, brain responses, and genetic variations in the context of reward sensitivity. Biological psychology **118**, 79–87 (2016)
12. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation **9**(8), 1735–1780 (1997)
13. Hyde, K.K., Novack, M.N., LaHaye, N., Parlett-Pelleriti, C., Anden, R., Dixon, D.R., Linstead, E.: Applications of supervised machine learning in autism spectrum disorder research: a review. Review Journal of Autism and Developmental Disorders **6**, 128–146 (2019)

14. Jack, A., Sullivan, C.A., Aylward, E., Bookheimer, S.Y., Dapretto, M., Gaab, N., Van Horn, J.D., Eilbott, J., Jacokes, Z., Torgerson, C.M., et al.: A neurogenetic analysis of female autism. Brain **144**(6), 1911–1926 (2021)
15. Jacquemont, S., Coe, B.P., Hersch, M., Duyzend, M.H., Krumm, N., Bergmann, S., Beckmann, J.S., Rosenfeld, J.A., Eichler, E.E.: A higher mutational burden in females supports a "female protective model" in neurodevelopmental disorders. The American Journal of Human Genetics **94**(3), 415–425 (2014)
16. Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M.: Fsl. Neuroimage **62**(2), 782–790 (2012)
17. Kaiser, M.D., Hudac, C.M., Shultz, S., Lee, S.M., Cheung, C., Berken, A.M., Deen, B., Pitskel, N.B., Sugrue, D.R., Voos, A.C.: Neural signatures of autism. Proceedings of the National Academy of Sciences **107**(49), 21223–21228 (2010)
18. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization (2017)
19. Lord, C., Brugha, T.S., Charman, T., Cusack, J., Dumas, G., Frazier, T., Jones, E.J., Jones, R.M., Pickles, A., State, M.W., et al.: Autism spectrum disorder. Nature reviews Disease primers **6**(1), 1–23 (2020)
20. Masi, A., DeMayo, M.M., Glozier, N., Guastella, A.J.: An overview of autism spectrum disorder, heterogeneity and treatment options. Neuroscience bulletin **33**, 183–193 (2017)
21. Parisot, S., Ktena, S.I., Ferrante, E., Lee, M., Moreno, R.G., Glocker, B., Rueckert, D.: Spectral graph convolutions for population-based disease prediction. In: Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20. pp. 177–185. Springer (2017)
22. Pinto, D., Delaby, E., Merico, D., Barbosa, M., Merikangas, A., Klei, L., Thiruvahindrapuram, B., Xu, X., Ziman, R., Wang, Z., et al.: Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. The American Journal of Human Genetics **94**(5), 677–694 (2014)
23. Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M.: Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. Neuroimage **15**(1), 273–289 (2002)
24. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: Neural Information Processing Systems (2017)
25. Vértes, P.E., Rittman, T., Whitaker, K.J., Romero-Garcia, R., Váša, F., Kitzbichler, M.G., Wagstyl, K., Fonagy, P., Dolan, R.J., Jones, P.B., et al.: Gene transcription profiles associated with inter-modular hubs and connection distance in human functional magnetic resonance imaging networks. Philosophical Transactions of the Royal Society B: Biological Sciences **371**(1705), 20150362 (2016)
26. Yang, H., Liu, J., Sui, J., Pearlson, G., Calhoun, V.D.: A hybrid machine learning method for fusing fmri and genetic data: combining both improves classification of schizophrenia. Frontiers in human neuroscience **4**, 192 (2010)