Words, Figures and Tables:

- 5946 Words
- 2 Figures
- 1 Table

Long title:

Towards 3D Deep Learning for neuropsychiatry: predicting Autism diagnosis using an interpretable Deep Learning pipeline applied to minimally processed structural MRI data

Short title:

3D Deep Learning on minimally processed brain structural MRI data for Autism prediction

AUTHORS: Mélanie Garcia[1,3], Clare Kelly[1,2,3]

AFFILIATIONS: [1]Department of Psychiatry at the School of Medicine, Trinity College Dublin, Dublin, Ireland, [2]School of Psychology, Trinity College Dublin, Dublin, Ireland, [3]Trinity College Institute of Neuroscience, Trinity College, Dublin, Ireland.

CORRESPONDING AUTHOR:

Clare Kelly,

Trinity College Institute of Neuroscience,

Trinity College,

Dublin 2,

Ireland

clare.kelly@tcd.ie

1- Abstract

By capitalizing on the power of multivariate analyses of large datasets, predictive modeling approaches are enabling progress toward robust and reproducible brain-based markers of neuropsychiatric conditions. While Deep Learning offers a particularly promising avenue to further advance progress, there are challenges related to implementation in 3D (best for MRI) and interpretability. Here, we address these challenges and describe an interpretable predictive pipeline for inferring Autism diagnosis using 3D Deep Learning applied to minimally processed structural MRI scans. We trained 3D Deep Learning models to predict Autism diagnosis using the openly available ABIDE I and II datasets (n = 1329, split into training, validation, and test sets). Importantly, we did not perform transformation to template space, to reduce bias and maximize sensitivity to structural alterations associated with Autism. Our models attained predictive accuracies equivalent to those of previous Machine Learning studies, while side-stepping the time- and resource-demanding requirement to first normalize data to a template, thus minimizing the time required to generate predictions. Further, our interpretation step, which identified brain regions that contributed most to accurate inference, revealed regional Autism-related alterations that were highly consistent with the literature, such as in a left-lateralized network of regions supporting language processing. We have openly shared our code and models to enable further progress towards remaining challenges, such as the clinical heterogeneity of Autism, and to enable the extension of our method to other neuropsychiatric conditions.

Abbreviations:

CNN: Convolutional Neural Networks, a category of Deep Learning algorithm

ML: Machine Learning

DL: Deep Learning

Med3dNet - Resnet50: pretrained Residual Networks model with 50 layers

DenseNet121: Densely Connected Convolutional Networks with 121 layers

Epoch: a hyperparameter that defines the number of times that the learning algorithm has optimized the parameters on the entire training dataset.

MRI: Magnetic Resonance Imaging

2- Introduction:

Autism Spectrum Disorder (Autism) is a complex and heterogeneous neurodevelopmental condition characterized by divergence from typical development on a number of behavioral dimensions, including communication, social interaction, and repetitive or restricted behaviors or areas of interest[1]. These manifest behaviors likely reflect developmental neurological alterations over the lifespan[2-7], a suggestion supported by structural MRI studies[2, 8-26]. Despite substantial research effort, however, no compelling brain-based biomarkers have yet emerged. Autism Spectrum Disorder is diagnosed through clinician judgment and gold standard observational tests, such as the Autism Diagnostic Observation Schedule (ADOS)[27] and the Autism Diagnostic Interview-Revised (ADI-R)[28], typically around age 43 months[29]. Given the considerable heterogeneity inherent to the diagnosis, and the wide range of long-term outcomes, the availability of robust and reproducible brain biomarkers for Autism could help refine diagnoses and treatment plans, thus promoting better outcomes. The availability of predictive models could also help clinicians build personalized care paths[30].

One challenge in the search for biomarkers and in the development of predictive models is the attainment of sample sizes that afford adequate statistical power. This challenge is exacerbated by clinical heterogeneity[30]. Multi-site collaborative studies yielding well-powered samples, such as ABIDE I and II[31-32], have gone some way to addressing this challenge, and analyses of these samples suggest a distributed pattern of Autism-related structural alterations[16-18, 20, 21, 24, 33-34]. The application of multivariate approaches, such as Machine and Deep Learning, offer another promising avenue for the search for brain-based biomarkers and the construction of predictive models.

These methods enable the simultaneous exploration of a very large set of features, offering much more powerful analytical capacity than univariate approaches. To date, such

approaches have had moderate success, with recently reported prediction accuracies (for Autism diagnosis) in the range of 65-70% for models built using both functional and structural MRI data[35-38]. In an effort to boost accuracy through competition, Traut et al.[39] held an international challenge in which competing teams predicted Autism diagnosis using a large multisite dataset comprising preprocessed anatomical and functional MRI data from > 2,000 individuals. Of the 589 models submitted, the 10 best were combined and evaluated using a subset of unseen data (from one of the sites included in the main dataset), as well as data from an additional, independent acquisition site. The blended model achieved an ROC AUC of ~0.66 using features extracted from anatomical data only. One observation from this effort was the fact that prediction accuracy increased with increasing sample size. Another was that while prediction accuracy for the subset of unseen data was similar to validation accuracy, accuracy for the novel site was poorer, illustrating the challenge of generalization, particularly to new data collection sites.

Although recent gains in prediction accuracy are promising, Machine Learning studies conducted to date have two main limitations. The first is that preprocessing pipelines often have many steps, each of which can introduce biases to prediction models. In particular, preprocessing typically includes transformation to a template space, such as MNI152, which was created using anatomical scans acquired from neurotypical adults. Template normalization may therefore negatively impact the ability to detect Autism-related alterations in brain structure, introduce biases, and lead to poorer reproducibility[30]. A second limitation is that datasets used for prediction tend to be clinically heterogeneous, but this heterogeneity is not explicitly accounted for in the models, leading to inconsistent results between separate datasets[40]. Many Autistic participants have a secondary diagnosis, which is often another psychological condition such as ADHD or anxiety, or a neurological condition such as epilepsy or Fragile X syndrome[17,24,26]. Ignoring these comorbidities may introduce biases or lead to non-specific biomarkers[17], since in such analyses, the label “autism” is not well delimited.

In the current study, we sought to develop a prediction pipeline that could overcome these challenges. To do this, we trained 3-Dimensional Deep Learning models to predict Autism diagnosis from minimally preprocessed structural MRI data, to avoid biases introduced by template normalization. To address the influence of clinical heterogeneity, we built our models using a large sample of 1329 patients (521 with autism) without comorbidities, following the classical framework of train-validate-test. To test if the patterns identified by the best models were robust to comorbidity, we tested the three best models on a second dataset comprising 270 patients (155 with autism) with comorbid diagnoses.

Deep Learning models can extract meaningful implicit features during the optimization process, which minimizes the preprocessing required and ultimately reduces prediction time. While 2D Deep Learning models are increasingly popular, 3D Deep Learning is not widely used in Medical Imaging applications, in part because of the large number of parameters to optimize (greater than in 2D) and concerns related to interpretability. To address the challenge of extracting information about predictive features (i.e., interpretability), we leveraged recently developed methods to build an interpretation pipeline that identifies predictive brain areas while avoiding the requirement for template normalization.

In this paper, we described our novel pipeline for interpretable 3D Deep Learning prediction of Autism diagnosis from structural MRI data. In our proof-of-concept analyses, our models achieved the same prediction accuracy as is typical for Machine Learning models, while avoiding the potential biases introduced by template normalization. Our interpretation pipeline identified a set of regions that replicated well across datasets (including participants with comorbidities), and models, and which converged with previous structural imaging studies on Autism. To facilitate further development of our pipeline, we have openly shared all our code through GitHub (https://github.com/garciaml/Autism-3D-CNN-brain-sMRI).

3- Materials and Methods:

3.1. Data and Quality Control

We used T1-weighted structural MRI data from the ABIDE I (980 scans) and II (857 scans) datasets[31-32] and 140 scans from ADHD200[41]. We performed quality control using BrainQCNet[42], retaining scans with a probability score below 60% as advised in [42]; 797 scans from ABIDE I, 704 from ABIDE II and 98 from ADHD200 remained after this step.

Our primary analysis focused on participants with a diagnosis of Autism but no reported comorbidity and comparison participants with no psychiatric diagnosis. Excluding participants with comorbidities resulted in a dataset of 1329 participants which were used for training, validating and testing the models.

All participants in the testing set (n = 65, 26 with Autism) were obtained from different (independent) data collection sites than participants in the training (n = 1074, 421 with Autism) and validation (n = 190, 74 with Autism) sets.

To examine the impact of comorbidities on prediction accuracy, we created a second evaluation set of participants who had at least other diagnosis in addition to Autism, such as ADHD, phobias, depression, and anxiety. This dataset (testing set 2) contained scans from 270 participants (155 with Autism diagnosis).

Further details on the datasets are provided in **S1 - Detailed Data Description.**

3.2. Preprocessing

We employed a minimal preprocessing pipeline that did not apply transformation to template space, to avoid any impact of brain normalization on the detecting of Autism-related alterations in brain structure. Instead, we applied FSL's Brain Extraction Tool (BET; https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/BET) to remove non-brain tissue, followed by a number of minor non-deforming transformations, to prepare our data to be processed by the Deep Learning algorithm:

- *Resolution homogenization:* the ABIDE datasets comprise data from different data collection sites, each of which has different scanners and acquisition protocols, Accordingly, the T1-weighted volumes have heterogeneous voxel spacing that could bias the analysis. We used Linear Interpolation to perform resampling, with the Resample function from the Python library TorchIO (https://torchio.readthedocs.io/_modules/torchio/transforms/preprocessing/spatial/resample.html#Resample), built from the Insight Toolkit (https://itk.org/Doxygen/html/index.html) to resample all volumes to a fixed resolution of 1.5mm*1.5mm*1.5mm. We also reordered the data to RAS+ orientation.
- *Intensity normalization:* We removed the noise generated by voxel value outliers in every image by truncating the intensities to the range of 0.5 to 99.5 percentiles using the RescaleIntensity function from TorchIO. We also normalized each volume by subtracting the mean intensity value v_m to each voxel value v_i , and then dividing by the standard deviation v_{sd} , obtaining a new voxel value v'_i .

$$v'_i = (v_i - v_m)/v_{sd}$$

- *Cropping or Padding:* We cropped or padded each volume to obtain a uniform shape for all the volumes of 256*256*256. This shape was sufficiently large to fit the full brains and was also appropriate as an input shape to our deep learning models, in view of the filters applied all along each network (described in detail below).

3.3. Classification Models in 3D

Comparing different types of algorithm enables the detection of overfitting and retention of the best type of algorithm for the given problem[43]. We compared two models: (1) DenseNet121[44] and (2) Med3D-ResNet50[45], well known Deep Learning algorithms with good 2D performance[44-45]. DenseNet121 is more compact and has fewer parameters than ResNet50 making it possible to train on 3D data, while Med3D-ResNet50 [45] is a version of ResNet50 that has been pre-trained on medical images, including brain sMRI scans. Logically, pre-trained models enable better convergence and performance on new data and tasks of the same context. We fine-tuned Med3d-ResNet50 to adapt it to our task by training the last convolutional layers (corresponding to the 4th convolutional block). We also appended the last classifier block, consisting of a global average pooling layer and a fully connected layer (see **S2 - Model architectures**)

Like in [44] and in [45], we used the ReLU function as the activation function, the cross entropy loss, and the Adam optimizer with a fixed learning rate of 0.001.

3.4. Interpreting outcomes of Deep Learning algorithms

3.4.1. Guided-Grad-CAM

In order to interpret and evaluate the reliability and relevance of our 3D Deep Learning models, we used Guided Grad-CAM[46], which combines guided backpropagation[47] and Grad-CAM[46]. This represents a good trade-off between the precision offered by feature maps produced by interpretability algorithms and the processing time required. Mathematically, guided Grad-CAM[46] is an element-wise product of the results of the two algorithms. It returns a high resolution map of the fine-grained features that is also class-discriminative.

In the context of our study, for a given trained CNN model (either DenseNet121 or Med3DNet-ResNet50), we used guided Grad-CAM to generate one “attention map” for each

participant at the inference step (i.e. the first layer of the CNN). This attention map matched the input scan resolution and voxel dimensions, and its voxel values corresponded to scores of "importance" for the prediction of Autism/non-Autism by the trained CNN model. Mathematically, for a given input participant's scan, we computed $q_{50\%}$ - the median of the voxel values of the attention map obtained with guided Grad-CAM. We then built a binary mask by returning all the voxel values lower than $q_{50\%}$ to 0 and all voxels greater than $q_{50\%}$ to 1. We used this mask M to identify the brain regions that are the most important for the prediction of Autism across the sample and across algorithms.

3.4.2. HighRes3DNet

As noted above, a key feature of our preprocessing pipeline was our avoidance of normalization to a group template. This creates a significant challenge for the identification of the brain areas that were most predictive of diagnosis across participants. We solved this challenge by segmenting individual scans into anatomical units and combining this information with the mask M created in the preceding step.

HighRes3DNet[48] is a Deep Learning algorithm that segments brain MRI scans following the GIF brain parcellation (V3, http://niftyweb.cs.ucl.ac.uk/program.php?p=GIF ; [49]). The GIF algorithm was especially built to be robust to brain morphological differences, especially those encountered in populations with atypical brain development like Autism[49].

We segmented each participant's brain with the HighRes3dNet algorithm (first homogenizing scans to voxel size 1mm*1mm*1mm using linear Interpolation). The resulting segmented images were resampled to 256*256*256 images of voxel size 1.5mm*1.5mm*1.5mm to match the resolution of the attention maps obtained from the guided Grad-CAM algorithm, while retaining the segmented voxel values.

Specifically, we know that the information on the transformations applied to the segmented image is contained into the affine matrix of the resulting transformed segmented image. Mathematically, we note $X = [x, y, z, 1]$, the column vector of the coordinates x, y, z of a voxel in a segmented image obtained with HighRes3DNet (voxel size: 1mm*1mm*1mm), $Y = [x', y', z', 1]$ the column vector of the coordinates x', y', z' of a voxel in the corresponding transformed segmented image (size: 256*256*256; voxel size: 1.5mm*1.5mm*1.5mm), and $A \in |R^4$ its affine matrix. We note B, the inverse matrix of A, such that $BA = A^{-1}A = I$, where I is the identity matrix in R^4.

Thus, we have the relationship:

$AX = Y$

$\Leftrightarrow X = BY$, $\forall (x', y', z') \in [1, 256]^3$.

Thus, if we take x', y', z' the coordinates of a voxel in the mask M obtained from guided Grad-CAM, we can obtain the corresponding x, y, z voxel coordinates in the segmented image, and thus get the voxel value and the name of the area at (x, y, z).

Applying this procedure for every scan, we obtained a table containing, for every area of the HighRes3DNet atlas, a relative frequency corresponding to the number of voxels in the area with value = 1, divided by the total number of voxels in this area in the segmented image. This relative frequency corresponds to the proportion of the area that is considered important for the prediction by a CNN model, for that participant. These proportions were then used to compare different brain areas and to draw up a ranking of brain areas for each model, dataset (training, validation, testing sets), and type of prediction (True Positives, True Negatives, False Positives, False Negatives), to improve interpretability for our CNN models.

3.5. Machine and Code availability

We trained our model on a GPU Nvidia RTX 3090 (24 GB memory) with a batch size of 2.

We openly shared the code of this project on GitHub, in the repository: https://github.com/garciaml/Autism-3D-CNN-brain-sMRI. The models are also shared so that they can be reused as pre-trained models for similar applications.

4. Results:

4.1. Training performance

For all the probability scores of all the models, we chose a threshold of 0.5 for the class "Autism diagnosis" to define the prediction and compute the accuracy and ROC AUC scores. We trained each model up to 100 epochs and computed model accuracy using the validation set (190 scans) every two epochs. Details on the validation set accuracy during training for the two models DenseNet161 and Med3d-ResNet50 are provided in **S3 Fig.1** in **S3 - Performance of the models.**

For ResNet50, the best validation set accuracy was 62.6%, achieved at 42 epochs. For DenseNet121, 66.3% accuracy was achieved at 32 epochs and 67.4% was achieved at 70 epochs. Next, we compared the performance of these three best models (one ResNet50 model and two DenseNet121 models) for the prediction of diagnosis in the training, validation, and testing sets.

4.2. Prediction Performance: Autism diagnosis

For the prediction of Autism diagnosis, the three best models behaved differently, as shown by the Receiver Operating Characteristic curves in **Fig 1**. Med3d-ResNet50-42ep overfitted the data - the accuracy and ROC AUC scores were very high on the training set (94.2% and 99.9% respectively) but much lower on the validation (acc = 62.6% and AUC = 62.1%) and testing sets (acc = 53.8% and AUC=57.3%). DenseNet121-32ep appeared to be more stable

in terms of its overall performance on the training (acc = 65.5% and AUC = 69.1%), validation (acc =66.3% and AUC = 68.8%) and testing (acc =55.4% and AUC = 60.7%) sets. DenseNet121-70ep had better performance on the training (acc = 69.7% and AUC = 77.1%) and validation (acc = 67.4% and AUC = 68.1%) sets than DenseNet121-32ep, but poorer performance on the testing set (acc = 40% and AUC = 38.1%).

Table 1 displays the sensitivity and specificity of each model for each dataset. DenseNet121-32ep exhibited high specificity on the training and validation sets, but low sensitivity. Paradoxically, it had high sensitivity but low specificity on the testing set. DenseNet121-70ep behaved similarly on the testing set while on the training and validation sets, sensitivity and specificity were balanced and fairly high. Finally, for Med3d-ResNet50-42ep, sensitivity and specificity were very high on the training set, unbalanced on the validation set with low sensitivity and very high specificity, and balanced on the testing set, but with moderate values.

Sensitivity on the second testing set, which included participants with comorbidities, was low for all models. This demonstrates that when the training and testing sets include only participants without known comorbidities, predicting Autism diagnosis for participants with comorbidities is particularly challenging. Here, we found that this produces a large increase in False Negatives in particular. One potential explanation is that neuroimaging markers of Autism are less salient when individuals have another diagnosis involving similar or other neuroimaging markers. Another explanation is that more data is needed to adequately train DL algorithms on the whole spectrum of Autism in the context of comorbidities.

Further details and comments on the performance of the models are given in **S3 - Performance of the models**, and a comparison of the predicted scores with the scores of diagnosis are given in **S4 - Analysis of ADI-R and ADOS scores, age, gender and full IQ.**

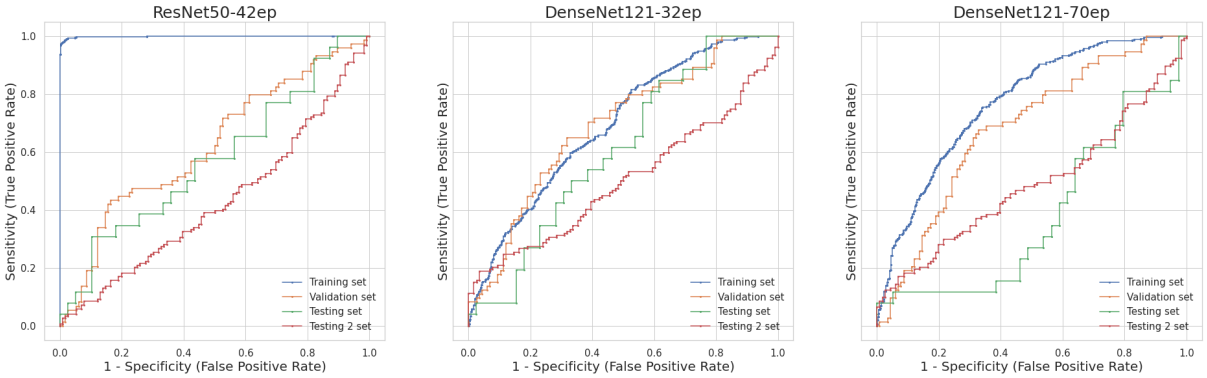


Fig 1. Receiver Operating Characteristic curves for all the three models and all the four datasets

	Med3dNet - Resnet50; trained on 42 epochs	DenseNet121; trained on 32 epochs	DenseNet121; trained on 70 epochs
Sensitivity	Train: 85,3%	Train: 32,8%	Train: 68,2%
	Validation: 17,6%	Validation: 36,5%	Validation: 66,2%
	Test: 50%	Test: 84,6%	Test: 69,2%
	Test 2: 8,4%	Test 2: 7,6%	Test 2: 31%
Specificity	Train: 100%	Train: 86,7%	Train: 70,8%
	Validation: 91,4%	Validation: 85,3%	Validation: 68,1%
	Test: 56,4%	Test: 35,9%	Test: 20,5%
	Test 2: 87,8%	Test 2: 100%	Test 2: 73%

Table 1. Sensitivity and Specificity of each model on each dataset (training, validation, testing sets with no comorbidity and testing set 2, which included patients with comorbidities).

4.3. Interpretability: *True Positive* discriminative ROIs

We segmented each participant's scan using HighRes3DNet (GIF parcellation), to extract a measure of "prediction importance" (the output of the guided Grad-CAM algorithm) for each of the three best models. We identified the regions that best contributed to True Positives

(TP), True Negatives (TN), False Positives (FP) and False Negatives (FN), across the whole dataset (training + validation + testing 1 & 2 sets).

For every pair of model and dataset, we defined the "most predictive" regions as those with relative frequency values (see **Section 3.6,** above) greater than the 90% percentile. This yielded 16 regions for each model and dataset pair. To compare the most predictive regions across models and datasets (training, validation, testing Set 1 - no comorbidities, testing set 2 - with comorbidities), we summed the presence (1) or absence (0) of the most predictive regions over all the datasets, separately for *True Positives* and *True Negatives*. Across all three models, 79 areas were found to be most predictive for *True Positives*, including 26 areas spanning both left and right hemisphere, 23 areas in the left hemisphere only, 3 areas in the right hemisphere only, and the Corpus Callosum. Retaining only areas that replicated across all four datasets (training, validation, and test 1/2), we found that areas in the left hemisphere were more replicable than those in the right, and that the majority of areas were in the prefrontal cortex. The **S5 - Most important regions for the prediction of True Positives** provides **S5 Table 6** that summarizes the most replicable regions across models and datasets that are important to predict *True Positives*, and a detailed analysis of these most replicable regions.

Overall, 17 regions were found to best predict *True Positives* across models and replicate across datasets (training, validation, and testing 1/2). These regions are shown in **Fig 2** and include regions in the left frontal lobe (medial frontal cortex, inferior and middle frontal gyrus, lateral and medial precentral gyrus, anterior and subcallosal cingulate gyrus, and posterior orbital gyrus), left temporal lobe (temporal pole, planum temporale, parahippocampal gyrus), parietal lobe (parietal operculum, supramarginal gyrus, and superior parietal lobe), as well as left parietal white matter and the right ventral thalamus.,

Looking at these data another way, and taking the regions that were most predictive across datasets and which replicated across the three models, we again obtained left hemisphere regions that are located in the frontal lobe - middle and inferior frontal gyrus (pars triangularis) and medial precentral gyrus - and in the limbic system and its associated structures - anterior cingulate gyrus, subgenual cingulate gyrus, parahippocampal gyrus (**Fig 2b**).

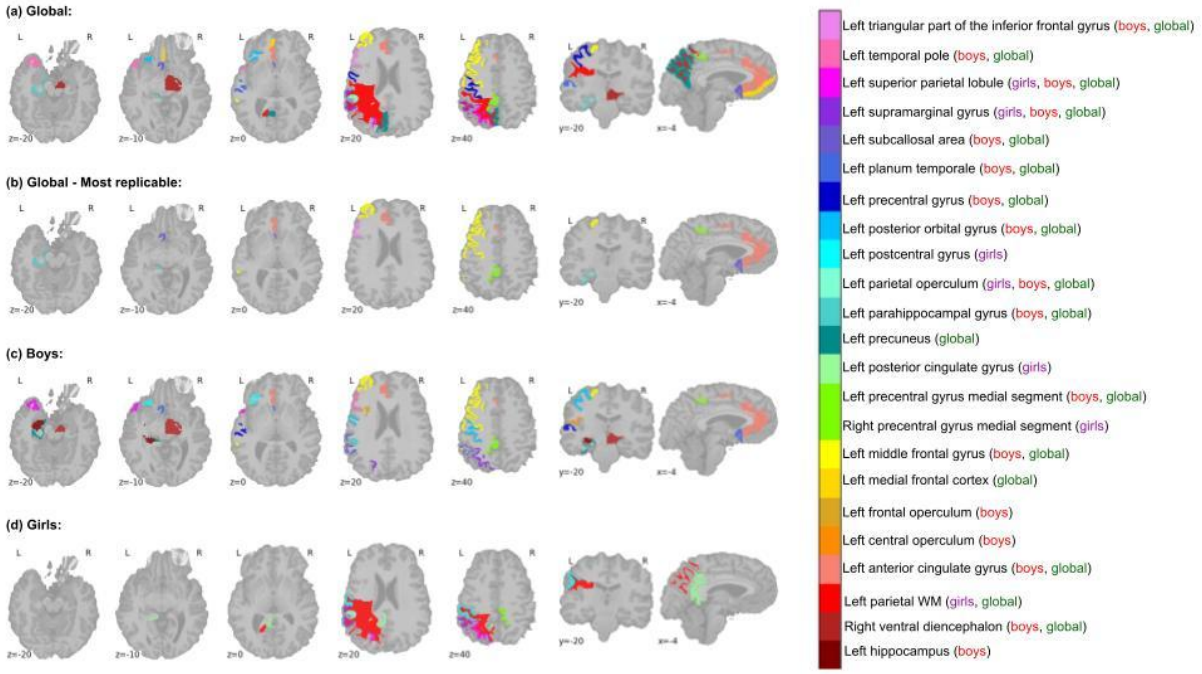


Fig 2. (a) Regions most predictive of Autism diagnosis; (b) Most predictive regions that replicate across datasets, (c) Most predictive regions for boys; (d) Most predictive regions for girls.

4.3.1. Effect of gender

Regions important for predicting *True Positives* for boys were different from those for girls. Regions common to both genders were located in the left parietal lobe: parietal operculum, supramarginal gyrus, and superior parietal lobule (**Fig 2c, d**). Globally, regions found important to predict *True Positives* for boys were more replicable across the datasets (training, validation, testing 1/2) than for girls. For boys, several left prefrontal regions were replicably predictive of Autism diagnosis: left anterior cingulate gyrus, middle frontal gyrus,

inferior frontal gyrus (pars triangularis; ResNet50-42ep only), medial precentral gyrus (DenseNet121-32ep) and precentral and parahippocampal gyrus (DenseNet121-70ep).

S8 - True Positives by Gender shows these results in **S8 Tables 10 and 11**.

4.3.2 Relationship with age

Autism has been associated with disrupted brain development across the lifespan. To assess whether there were any developmental trends in the most predictive areas, we created four age categories (5-10yrs, 10-15yrs, 15-20yrs, >20yrs) and identified the most predictive (True Positives) regions for each category, separately for boys and girls. **S 9 - True Positives by Gender and Age** shows these results.

Our results showed that the most discriminative regions varied with the age. In particular, left precentral gyrus, central operculum, and posterior orbital gyrus replicably predicted *True Positives* in boys aged 5-10yrs, while left inferior frontal gyrus (pars triangularis), subcallosal/subgenual cingulate cortex, and supramarginal gyrus, were most predictive for boys aged 10-15 years old.

In addition, we found that the replicability of each region decreased as age increased. Indeed, we found that the left and inferior frontal (pars triangularis) gyrus, posterior orbital gyrus and putamen were most predictive for 15-20 years old, but only for participants without comorbidities. Left temporal areas - parahippocampal gyrus, superior temporal gyrus and temporal pole - were most predictive for males aged 20-64yrs without comorbidity.

Examining global prediction performance for these different age groups reveals other interesting trends, such as a decrease in the number of False Negatives and True Negatives with increasing age, for both boys and girls. This suggests that our prediction of Autism diagnosis tended to be more sensitive but less specific as age increased.

4.3.3. True Negatives

We adopted the same approach described above to identify regions most predictive of *True Negatives* (i.e., absence of an Autism diagnosis). The results (see **S6 - Most important regions for the prediction of True Negatives**) showed that the most replicable regions for predicting *True Negatives* were in the left hemisphere and included the frontal operculum, the precuneus, the planum polare, the inferior occipital gyrus, the occipital fusiform gyrus, the superior occipital gyrus and the thalamus proper. It also included the cerebellar vermal lobules VI and VII.

Another result is that the regions left precuneus, parietal operculum, and superior parietal lobe, and right thalamus were important (at various degrees of replicability and for different models) for the prediction of both *True Negatives* and *True Positives*. The 23 other regions important for the prediction of True Negatives are different from those that were important to the prediction of True Positives.

4.3.4. Bad predictions - False Positives and False Negatives

We adopted the same approach described above to identify regions most predictive of *False Positives* (i.e., incorrectly predicted Autism diagnosis) and *False Negatives* (i.e., incorrectly failed to predict Autism diagnosis). **S7- Most replicable regions for *False Positives* and *False Negatives*** shows these results. No highly replicable regions (replicable over all datasets) were found for *False Positives.* However, regions with a high level of replicability for *False Positives* for DenseNet121-70 overlapped with replicable regions for the prediction of *True Positives* for the two other models and included the middle frontal gyrus, precentral gyrus medial segment, and triangular part of inferior frontal gyrus. This illustrates differences in the calibration of each algorithm and demonstrates the importance of comparing different models. For *False Negatives*, the most replicable regions were again found in the left

hemisphere and included the left frontal operculum, left precuneus, left superior temporal gyrus, left planum polare, left inferior occipital gyrus and left occipital fusiform gyrus.

4.4. Does image background contribute to model predictions?

As a final test, we examined whether image background (i.e., information outside the brain) contributed to predictions. For Med3d-ResNet50-42ep the relative frequency of the Background (RF) is the smallest (RF=0.97%) and the second smallest for DenseNet121-70ep (RF=0.28%), meaning that this area is not considered predictive for the models. For DenseNet121-32ep, it is among the last 4% informative areas of the model (RF=0.74%). These results confirm that the models use information from inside rather than outside the brain to make a prediction, supporting their validity.

4.5. Multi-site effect

We observed an inhomogeneous consistency of the distributions of probability scores between the different sites (see **S10 - Multi-site effect**). We displayed the accuracy scores for every site in the whole dataset (training+validation+testing sets) in **S10 Table 20**, and it also confirmed the multi-site effect.

5. Discussion:

This study outlines and demonstrates a novel approach for inferring Autism diagnosis from structural brain imaging data using 3D Deep Learning algorithms. To maximize the interpretability of the model outputs, we also used a second type of algorithm - guided Grad-CAM[46] - to extract patterns important for the predictions. This step revealed a set of regions predominantly located in the left hemisphere, including lateral and medial prefrontal cortex, anterior cingulate, the superior temporal gyrus, lateral parietal regions including supramarginal gyrus, parahippocampal gyrus. The only right hemisphere region highlighted in our analyses was the right thalamus. The regions highlighted by this interpretability

analysis, the brain structural features of which were most important for accurate inference of Autism diagnosis (i.e., *True Positives*), are highly consistent with the literature. Our predictive modeling framework has considerable potential to be extended to further datasets to identify and refine sensitive and specific brain biomarkers of Autism using MRI data.

5.1. 3D Deep Learning applied to minimally processed data

To our knowledge, this is the first time that 3D-DL CNNs have been used to predict Autism diagnosis from 3D structural MRI scans. Our findings show that these algorithms are capable of inferring Autism diagnosis on the basis of structural MRIs with at least the same level of accuracy as traditional Machine Learning algorithms, while requiring a smaller number of training epochs. The average accuracy score (64.1%) and ROC AUC score (0.67) obtained for participants without comorbidities is consistent with previous Machine Learning models trained on sMRI data (e.g., [39]). The comparable accuracy we achieved should be viewed in the context of the speed of inference of Deep Learning models over Machine Learning approaches. While Machine Learning algorithms require inputs derived following extensive preprocessing of structural MRI data, including normalization to template space, our Deep Learning models used minimally preprocessed data. In particular, we avoided transformation to template space, a near-universal requirement of neuroimaging analyses that may negatively impact the ability to detect structural alterations associated with the diagnosis of interest. Although our pipeline included some minimal preprocessing steps to address the fact that a diversity of scanners and acquisition protocols was used across data collection sites, resulting in heterogeneous voxel spacing and signal intensities. Resolution homogenization and intensity normalization were applied to address these variations, and it is possible that these steps could bias the algorithm. Further, despite these steps, a clear effect of the data collection site was observable. Future studies will incorporate specific preprocessing steps like the ComBat algorithm[50] to integrate scan parameters during training and minimize site effects.

5.2. Interpretability

The outputs of Deep Learning models are not straightforwardly understandable, giving rise to the challenge of poor interpretability. This challenge arises because mathematically, Deep Learning models are composed of multiple functions. Each of these functions is nonlinear and is itself the sum of multiple functions. Further, models such as the 3D CNNs used in the current study have a large number of parameters that must be optimized. One of the goals of our study was to address this drawback by devising a pipeline that would allow for the extraction of predictive brain regions, providing interpretability. Guided Grad-CAM[46] was chosen for this purpose, due to its reasonable computation time and its ability to return fine-grained class-specific segmentations of important (predictive) voxels in the input images.

A challenge for our novel interpretability process was to identify brain areas that were predictive of Autism diagnosis across participants while avoiding the requirement for template normalization. To address this issue, we used a segmentation algorithm to partition individual volumes into established anatomical regions. We used HighRes3DNet[48] for this task because it was built to be pathology-agnostic, robust to brain morphology differences, and has reduced computation time compared to other algorithms (e.g., the GIF algorithm[49]). We performed a detailed analysis of the regions that were most relevant for inferring an Autism diagnosis, by examining true and false positives and negatives separately for each dataset and algorithm. We also identified regions that were reproducibly identified across algorithms and datasets. This detailed analysis is important because each model has biases, likely resulting in a differential weighting of anatomical features and brain areas. This analysis showed that regions of left prefrontal cortex (inferior and middle frontal gyrus, medial prefrontal gyrus, anterior and subgenual cingulate cortex), along with the parahippocampal gyrus were brain regions whose morphological features contributed most to the accurate inference of Autism across models and datasets without and with comorbidities. The areas highlighted are consistent with previous studies reporting

Autism-related disruptions to cortical development[24, 51-53] and gyrification processes[24, 54] in these regions. Further, also consistent with the literature, we found that the most predictive regions varied according to both gender and age, as well as the presence of comorbidities[17, 24, 55]. This is consistent with observations that Autism is a complex condition, with patterns of neurological divergence that vary with age[17, 24, 52-53] and sex[17, 24, 55].

Reproducibly predictive regions in the limbic system (left parahippocampal gyrus, anterior cingulate gyrus, and subcallosal area), dorsal medial frontal cortex, and precentral gyrus fit well with previous work on the role of atypical socio-emotional and motor circuitry in Autism[56-60]. Many of the left-hemisphere regions identified as contributing to accurate inference of Autism diagnosis fall within the canonical left-lateralized language network, including inferior prefrontal and inferior parietal regions, and the planum temporale in superior temporal gyrus[61-63]. Divergent structure and function in the language network is a robust and reproducible finding in Autism[64-67]. Since early language processing appears to be an important predictor of long-term outcomes in Autism[68-70] identification of early-emerging structural alterations in the underlying language network has the potential to yield a powerful marker of Autism or Autism subtypes, which could, in turn, direct individualized interventions and improve prognosis.

An important caveat is that while our novel interpretation step identified which regions of the brain had morphological features relevant to the model-based inference of Autism, it did not provide information on what these morphological features were. For example, features such as cortical thickness, the location of the gray-white boundary, surface area, and gyral/sulcal morphometry could all play a role in prediction of Autism[22,52,71]; and different morphological features may be relevant in different brain areas. While the precise nature of the Autism-related morphological features are not discernable from our analyses, our predictive modeling analyses can be followed up with in-depth, targeted, and

hypothesis-driven examinations of the areas highlighted in independent samples to uncover the nature of these features.

5.3. Limitations

Our pipeline for prediction of neuropsychiatric diagnosis (Autism) on the basis of minimally preprocessed T1 MRI scans advances progress toward interpretable 3D Deep Learning applications in biological psychiatry and toward the identification of reproducible brain biomarkers that will help refine diagnoses and treatment plans across conditions. Our study had several limitations, however, which may be addressed in further refinements of our pipeline.

First, we trained our models on 100 epochs, which is an acceptable number relative to other studies using 3D MRI scans[72], but which may have limited the convergence and optimization of the algorithms. Future work may train on a larger number of epochs or may employ earlystopping[73] to optimize training. Using the entire structural MRI scans (to explore prediction across the whole brain) may also have posed a challenge for convergence towards the “True” solution. Further, although we used a large dataset (1074 participants to train the models, 525 to validate and test the models), the amount of data available is still rather limited when we consider the clinical heterogeneity of Autism. This idea is supported by the poor prediction performance we observed for test set 2, which included participants with comorbid diagnoses (average accuracy = 46.3%, ROC AUC = 0.47 and average sensitivity = 15.7%). There are still questions in the literature about whether predicting a binary label, “Autism vs non-Autism” is a useful or appropriate endeavor, since Autism is a wide spectrum of behaviors and abilities which may encompass as many as four subtypes[23], and there is also considerable overlap of symptoms and neuromarkers across psychological conditions[17]. Future analyses will need to leverage even larger datasets to better address the clinical heterogeneity of Autism and to explore the prediction of categories beyond Autism and non-Autism.

Another limitation is related to the segmentation algorithm we used in the interpretation step. We used HighRes3DNet[48] to obtain rapid segmentation for each brain using the GIF algorithm[49], which was built to be robust on atypically developing brains. The segmentation produced is rather coarse, however - the algorithm outputs relatively large parcels, encompassing anatomically heterogeneous regions such as the anterior cingulate gyrus or superior parietal lobule. Further, as noted above, while our interpretation process localized regions that were important for prediction of Autism, it did not provide information on what the predictive morphological features of those regions were.

5.4. Future Directions

There is considerable scope to extend our interpretable Deep Learning pipeline to the prediction of other neurological or neuropsychiatric conditions or to other MRI modalities. Traut et al.[39] reported that prediction of Autism was considerably improved (from AUC=0.66 using only anatomical MRI to AUC=0.79 using both anatomical and functional data) for a blended model that incorporated both functional and structural MRI data. Future work will examine whether functional MRI data can also improve our models. Other efforts to improve our model will include training the models on more epochs, exploring other architectures, integrating scanning parameters and other confounds such as gender and age, and using different and extended class labeling. We have shared all our code (https://github.com/garciaml/Autism-3D-CNN-brain-sMRI) to enable other researchers to apply, reuse, and further develop our models and approach.

6. Conclusion

In this paper, we described a novel methodology to build a predictive model to infer Autism diagnosis using 3D Deep Learning applied to structural MRI scans, coupled with an interpretation step in the form of a descriptive method that identified the brain regions that

were most important for accurate inference. Importantly, we applied our models to minimally preprocessed data - completely avoiding the template normalization step, which may obscure diagnosis-related alterations in brain structure. We found that the predictive performance of our models was equivalent to that of Machine Learning models reported in the literature, while requiring less time to generate predictions (due to minimal preprocessing). There is considerable scope to refine our method or to incorporate other modalities (e.g., fMRI) to further boost predictive performance.

Our method for interpreting the output of Deep Learning models revealed highly predictive brain regions that were consistent with the literature, demonstrating that 3D Deep Learning models produce biologically plausible results without a priori knowledge or the requirement for pre-computation of morphological derivatives (e.g., volumes, cortical thickness, surface area). Although challenges related to the clinical heterogeneity of Autism remain to be addressed, we have openly shared our code and models for others to build on and extend, and to further progress the field towards the identification of robust and reproducible brain biomarkers for neuropsychiatric conditions.

7. Acknowledgements

We have much gratitude and appreciation for the Irish Research Council and the Hypercube Institute (Paris) for their funding.

We also thank Pr. Louise Gallagher as well as Dr. Robert Whelan for their feedback and guidance.

8. Disclosure of competing interests

None.

10. References

1. APA, Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). American Psychiatric Association Publishing; 2013.

2. Fishman I, Linke AC, Hau J, Carper RA, Müller R-A. Atypical Functional Connectivity of Amygdala Related to Reduced Symptom Severity in Children With Autism. Journal of the American Academy of Child & Adolescent Psychiatry. 2018;57: 764-774.e3. doi:10.1016/j.jaac.2018.06.015

3. McKinnon CJ, Eggebrecht AT, Todorov A, Wolff JJ, Elison JT, Adams CM, et al. Restricted and Repetitive Behavior and Brain Functional Connectivity in Infants at Risk for Developing Autism Spectrum Disorder. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging. 2019;4: 50–61. doi:10.1016/j.bpsc.2018.09.008

4. Lake EMR, Finn ES, Noble SM, Vanderwal T, Shen X, Rosenberg MD, et al. The Functional Brain Organization of an Individual Allows Prediction of Measures of Social Abilities Transdiagnostically in Autism and Attention-Deficit/Hyperactivity Disorder. Biological Psychiatry. 2019;86: 315–326. doi:10.1016/j.biopsych.2019.02.019

5. Walbrin J, Downing P, Koldewyn K. Neural responses to visually observed social interactions. Neuropsychologia. 2018;112: 31–39. doi:10.1016/j.neuropsychologia.2018.02.023

6. Jiang R, Calhoun VD, Zuo N, Lin D, Li J, Fan L, et al. Connectome-based individualized prediction of temperament trait scores. NeuroImage. 2018;183: 366–374. doi:10.1016/j.neuroimage.2018.08.038

7. Baker JT, Dillon DG, Patrick LM, Roffman JL, Brady RO, Pizzagalli DA, et al. Functional connectomics of affective and psychotic pathology. Proceedings of the National Academy of Sciences. 2019;116: 9050–9059. doi:10.1073/pnas.1820780116

8. Emerson RW, Adams C, Nishino T, Hazlett HC, Wolff JJ, Zwaigenbaum L, et al. Functional neuroimaging of high-risk 6-month-old infants predicts a diagnosis of autism at 24 months of age. Sci Transl Med. 2017;9: eaag2882. doi:10.1126/scitranslmed.aag2882

9. Subbaraju V, Suresh MB, Sundaram S, Narasimhan S. Identifying differences in brain activities and an accurate detection of autism spectrum disorder using resting state functional-magnetic resonance imaging : A spatial filtering approach. Medical Image Analysis. 2017;35: 375–389. doi:10.1016/j.media.2016.08.003

10. Heinsfeld AS, Franco AR, Craddock RC, Buchweitz A, Meneguzzi F. Identification of autism spectrum disorder using deep learning and the ABIDE dataset. NeuroImage: Clinical. 2018;17: 16–23. doi:10.1016/j.nicl.2017.08.017

11. Dickie EW, Ameis SH, Shahab S, Calarco N, Smith DE, Miranda D, et al. Personalized Intrinsic Network Topography Mapping and Functional Connectivity Deficits in Autism Spectrum Disorder. Biological Psychiatry. 2018;84: 278–286. doi:10.1016/j.biopsych.2018.02.1174

12. McKinnon CJ, Eggebrecht AT, Todorov A, Wolff JJ, Elison JT, Adams CM, et al. Restricted and Repetitive Behavior and Brain Functional Connectivity in Infants at Risk for Developing Autism Spectrum Disorder. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging. 2019;4: 50–61. doi:10.1016/j.bpsc.2018.09.008

13. Lake EMR, Finn ES, Noble SM, Vanderwal T, Shen X, Rosenberg MD, et al. The Functional Brain Organization of an Individual Allows Prediction of Measures of Social Abilities Transdiagnostically in Autism and Attention-Deficit/Hyperactivity Disorder. Biological Psychiatry. 2019;86: 315–326. doi:10.1016/j.biopsych.2019.02.019

14. Kishida KT, De Asis-Cruz J, Treadwell-Deering D, Liebenow B, Beauchamp MS, Montague PR. Diminished single-stimulus response in vmPFC to favorite people in children diagnosed with Autism Spectrum Disorder. Biological Psychology. 2019;145: 174–184. doi:10.1016/j.biopsycho.2019.04.009

15. Riddle K, Cascio CJ, Woodward ND. Brain structure in autism: a voxel-based morphometry analysis of the Autism Brain Imaging Database Exchange (ABIDE). Brain Imaging and Behavior. 2017;11: 541–551. doi:10.1007/s11682-016-9534-5

16. Ha S, Sohn I-J, Kim N, Sim HJ, Cheon K-A. Characteristics of Brains in Autism Spectrum Disorder: Structure, Function and Connectivity across the Lifespan. Exp Neurobiol. 2015;24: 273–284. doi:10.5607/en.2015.24.4.273

17. Ecker C, Bookheimer SY, Murphy DGM. Neuroimaging in autism spectrum disorder: brain structure and function across the lifespan. Lancet Neurol. 2015;14: 1121–1134. doi:10.1016/S1474-4422(15)00050-2

18. Yang DY-J, Beam D, Pelphrey KA, Abdullahi S, Jou RJ. Cortical morphological markers in children with autism: a structural magnetic resonance imaging study of thickness, area, volume, and gyrification. Mol Autism. 2016;7: 11. doi:10.1186/s13229-016-0076-x

19. Haar S, Berman S, Behrmann M, Dinstein I. Anatomical Abnormalities in Autism? Cereb Cortex. 2016;26: 1440–1452. doi:10.1093/cercor/bhu242

20. Pereira AM, Campos BM, Coan AC, Pegoraro LF, de Rezende TJR, Obeso I, et al. Differences in Cortical Structure and Functional MRI Connectivity in High Functioning Autism. Frontiers in Neurology. 2018;9. Available: https://www.frontiersin.org/article/10.3389/fneur.2018.00539

21. Bedford SA, Park MTM, Devenyi GA, Tullo S, Germann J, Patel R, et al. Large-scale analyses of the relationship between sex, age and intelligence quotient heterogeneity and cortical morphometry in autism spectrum disorder. Mol Psychiatry. 2020;25: 614–628. doi:10.1038/s41380-019-0420-6

22. Hong S-J, Valk SL, Di Martino A, Milham MP, Bernhardt BC. Multidimensional Neuroanatomical Subtyping of Autism Spectrum Disorder. Cereb Cortex. 2018;28: 3578–3588. doi:10.1093/cercor/bhx229

23. Hong S-J, Vogelstein JT, Gozzi A, Bernhardt BC, Yeo BTT, Milham MP, et al. Toward Neurosubtypes in Autism. Biological Psychiatry. 2020;88: 111–128. doi:10.1016/j.biopsych.2020.03.022

24. Pagnozzi AM, Conti E, Calderoni S, Fripp J, Rose SE. A systematic review of structural MRI biomarkers in autism spectrum disorder: A machine learning perspective. Int J Dev Neurosci. 2018;71: 68–82. doi:10.1016/j.ijdevneu.2018.08.010

25. Zheng W, Zhao Z, Zhang Z, Liu T, Zhang Y, Fan J, et al. Developmental pattern of the cortical topology in high-functioning individuals with autism spectrum disorder. Hum Brain Mapp. 2020;42: 660–675. doi:10.1002/hbm.25251

26. Sha Z, Wager TD, Mechelli A, He Y. Common Dysfunction of Large-Scale Neurocognitive Networks Across Psychiatric Disorders. Biological Psychiatry. 2019;85: 379–388. doi:10.1016/j.biopsych.2018.11.011

27. Lord C, Rutter M, Goode S, Heemsbergen J, Jordan H, Mawhood L, et al. Austism diagnostic observation schedule: A standardized observation of communicative and social behavior. J Autism Dev Disord. 1989;19: 185–212. doi:10.1007/BF02211841

28. Lord C, Rutter M, Le Couteur A. Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord. 1994;24: 659–685. doi:10.1007/BF02172145

29. 1. van 't Hof M, Tisseur C, van Berckelear-Onnes I, van Nieuwenhuyzen A, Daniels AM, Deen M, et al. Age at autism spectrum disorder diagnosis: A systematic review and meta-analysis from 2012 to 2019. Autism. 2021;25: 862–873. doi:10.1177/1362361320971107

30. Horien C, Floris DL, Greene AS, Noble S, Rolison M, Tejavibulya L, et al. Functional Connectome–Based Predictive Modeling in Autism. Biological Psychiatry. 2022 [cited 15 Sep 2022]. doi:10.1016/j.biopsych.2022.04.008

31. Di Martino A, Yan C-G, Li Q, Denio E, Castellanos FX, Alaerts K, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol Psychiatry. 2014;19: 659–667. doi:10.1038/mp.2013.78

32. Di Martino A, O'Connor D, Chen B, Alaerts K, Anderson JS, Assaf M, et al. Enhancing studies of the connectome in autism using the autism brain imaging data exchange II. Sci Data. 2017;4: 170010. doi:10.1038/sdata.2017.10

33. Zhang W, Ma L, Yang M, Shao Q, Xu J, Lu Z, et al. Cerebral organoid and mouse models reveal a RAB39b–PI3K–mTOR pathway-dependent dysregulation of cortical development leading to macrocephaly/autism phenotypes. Genes Dev. 2020;34: 580–597. doi:10.1101/gad.332494.119

34. Nakagawa N, Plestant C, Yabuno-Nakagawa K, Li J, Lee J, Huang C-W, et al. Memo1-Mediated Tiling of Radial Glial Cells Facilitates Cerebral Cortical Development. Neuron. 2019;103: 836-852.e5. doi:10.1016/j.neuron.2019.05.049

35. Wang Y, Wang J, Wu F-X, Hayrat R, Liu J. AIMAFE: Autism spectrum disorder identification with multi-atlas deep feature representation and ensemble learning. J Neurosci Methods. 2020;343: 108840. doi:10.1016/j.jneumeth.2020.108840

36. Lu H, Liu S, Wei H, Tu J. Multi-kernel fuzzy clustering based on auto-encoder for fMRI functional network. Expert Systems with Applications. 2020;159: 113513. doi:10.1016/j.eswa.2020.113513

37. Arya D, Olij R, Gupta DK. Fusing Structural and Functional MRIs using Graph Convolutional Networks for Autism Classification. : 18.

38. Dekhil O, Ali M, Haweel R, Elnakib Y, Ghazal M, Hajjdiab H, et al. A Comprehensive Framework for Differentiating Autism Spectrum Disorder From Neurotypicals by Fusing Structural MRI and Resting State Functional MRI. Semin Pediatr Neurol. 2020;34: 100805. doi:10.1016/j.spen.2020.100805

39. Traut N, Heuer K, Lemaître G, Beggiato A, Germanaud D, Elmaleh M, et al. Insights from an autism imaging biomarker challenge: promises and threats to biomarker discovery. Radiology and Imaging; 2021 Nov. doi:10.1101/2021.11.24.21266768

40. Benkarim O, Paquola C, Park B, Kebets V, Hong S-J, Wael RV de, et al. Population heterogeneity in clinical cohorts affects the predictive accuracy of brain imaging. PLOS Biology. 2022;20: e3001627. doi:10.1371/journal.pbio.3001627

41. Bellec P, Chu C, Chouinard-Decorte F, Benhajali Y, Margulies DS, Craddock RC. The Neuro Bureau ADHD-200 Preprocessed repository. NeuroImage. 2017;144: 275–286. doi:10.1016/j.neuroimage.2016.06.034

42. Garcia M, Dosenbach N, Kelly C. BrainQCNet: a Deep Learning attention-based model for multi-scale detection of artifacts in brain structural MRI scans. bioRxiv; 2022. p. 2022.03.11.483983. doi:10.1101/2022.03.11.483983

43. Hastie T, Friedman J, Tibshirani R. Introduction. In: Hastie T, Friedman J, Tibshirani R, editors. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York, NY: Springer; 2001. pp. 1–8. doi:10.1007/978-0-387-21606-5_1

44. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. 2017. pp. 4700–4708. Available: https://openaccess.thecvf.com/content_cvpr_2017/html/Huang_Densely_Connected_Convolutional_CVPR_2017_paper.html

45. Chen S, Ma K, Zheng Y. Med3D: Transfer Learning for 3D Medical Image Analysis. arXiv:190400625 [cs]. 2019 [cited 21 Mar 2022]. Available: http://arxiv.org/abs/1904.00625

46. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. arXiv:161002391 [cs]. 2019 [cited 19 Mar 2022]. doi:10.1007/s11263-019-01228-7

47. Springenberg JT, Dosovitskiy A, Brox T, Riedmiller M. Striving for Simplicity: The All Convolutional Net. arXiv:14126806 [cs]. 2015 [cited 19 Mar 2022]. Available: http://arxiv.org/abs/1412.6806

48. Li W, Wang G, Fidon L, Ourselin S, Cardoso MJ, Vercauteren T. On the Compactness, Efficiency, and Representation of 3D Convolutional Networks: Brain Parcellation as a Pretext Task. In: Niethammer M, Styner M, Aylward S, Zhu H, Oguz I, Yap P-T, et al., editors. Information Processing in Medical Imaging. Cham: Springer International Publishing; 2017. pp. 348–360. doi:10.1007/978-3-319-59050-9_28

49. Cardoso MJ, Modat M, Wolz R, Melbourne A, Cash D, Rueckert D, et al. Geodesic Information Flows: Spatially-Variant Graphs and Their Application to Segmentation and Fusion. IEEE Transactions on Medical Imaging. 2015;34: 1976–1988. doi:10.1109/TMI.2015.2418298

50. Radua J, Vieta E, Shinohara R, Kochunov P, Quidé Y, Green MJ, et al. Increased power by harmonizing structural MRI site differences with the ComBat batch adjustment method in ENIGMA. NeuroImage. 2020;218: 116956. doi:10.1016/j.neuroimage.2020.116956

51. Nordahl CW, Dierker D, Mostafavi I, Schumann CM, Rivera SM, Amaral DG, et al. Cortical folding abnormalities in autism revealed by surface-based morphometry. J Neurosci. 2007;27: 11725–11735. doi:10.1523/JNEUROSCI.0777-07.2007

52. Zielinski BA, Prigge MBD, Nielsen JA, Froehlich AL, Abildskov TJ, Anderson JS, et al. Longitudinal changes in cortical thickness in autism and typical development. Brain. 2014;137: 1799–1812. doi:10.1093/brain/awu083

53. Chien Y-L, Chen Y-C, Chiu Y-N, Tsai W-C, Gau SS-F. A translational exploration of the effects of WNT2 variants on altered cortical structures in autism spectrum disorder. J Psychiatry Neurosci. 2021;46: E647–E658. doi:10.1503/jpn.210022

54. Kohli JS, Kinnear MK, Fong CH, Fishman I, Carper RA, Müller R-A. Local Cortical Gyrification is Increased in Children With Autism Spectrum Disorders, but Decreases Rapidly in Adolescents. Cereb Cortex. 2019;29: 2412–2423. doi:10.1093/cercor/bhy111

55. Retico A, Giuliano A, Tancredi R, Cosenza A, Apicella F, Narzisi A, et al. The effect of gender on the neuroanatomy of children with autism spectrum disorders: a support vector machine case-control study. Mol Autism. 2016;7: 5. doi:10.1186/s13229-015-0067-3

56. Ameis SH, Catani M. Altered white matter connectivity as a neural substrate for social impairment in Autism Spectrum Disorder. Cortex. 2015;62: 158–181. doi:10.1016/j.cortex.2014.10.014

57. Mundy P. Annotation: The neural basis of social impairments in autism: the role of the dorsal medial-frontal cortex and anterior cingulate system - Mundy - 2003 - Journal of Child Psychology and Psychiatry - Wiley Online Library. [cited 18 Mar 2022]. Available: https://acamh.onlinelibrary.wiley.com/doi/abs/10.1111/1469-7610.00165?casa_token=svx2GLlhu0EAAAAA%3ACmkXUsvSjxZCIp3nRA5RHxhWcCvJVaPsfAtY5phXVYfg6qZVRcUhVrTuzJfAE2UYXALON2qJjnDuFJDI

58. Patriquin MA, DeRamus T, Libero LE, Laird A, Kana RK. Neuroanatomical and neurofunctional markers of social cognition in autism spectrum disorder. Hum Brain Mapp. 2016;37: 3957–3978. doi:10.1002/hbm.23288

59. Nebel MB, Eloyan A, Barber AD, Mostofsky SH. Precentral gyrus functional connectivity signatures of autism. Front Syst Neurosci. 2014;8: 80. doi:10.3389/fnsys.2014.00080

60. Carper RA, Solders S, Treiber JM, Fishman I, Müller R-A. Corticospinal Tract Anatomy and Functional Connectivity of Primary Motor Cortex in Autism. Journal of the American Academy of Child & Adolescent Psychiatry. 2015;54: 859–867. doi:10.1016/j.jaac.2015.07.007

61. Kelly C, Uddin LQ, Shehzad Z, Margulies DS, Castellanos FX, Milham MP, et al. Broca's region: linking human brain functional connectivity data and non-human primate tracing anatomy studies. Eur J Neurosci. 2010;32: 383–398. doi:10.1111/j.1460-9568.2010.07279.x

62. Malik-Moraleda S, Ayyash D, Gallée J, Affourtit J, Hoffmann M, Mineroff Z, et al. An investigation across 45 languages and 12 language families reveals a universal language network. Nat Neurosci. 2022;25: 1014–1019. doi:10.1038/s41593-022-01114-5

63. McAvoy M, Mitra A, Coalson RS, d'Avossa G, Keidel JL, Petersen SE, et al. Unmasking Language Lateralization in Human Brain Intrinsic Activity. Cereb Cortex. 2016;26: 1733–1746. doi:10.1093/cercor/bhv007

64. Floris, D. L., Lai, M. C., Auer, T., Lombardo, M. V., Ecker, C., Chakrabarti, B., ... & Suckling, J. (2016). Atypically rightward cerebral asymmetry in male adults with autism stratifies individuals with and without language delay. Human brain mapping, 37(1), 230-253.

65. Lindell AK, Hudry K. Atypicalities in cortical structure, handedness, and functional lateralization for language in autism spectrum disorders. Neuropsychol Rev. 2013;23: 257–270. doi:10.1007/s11065-013-9234-5

66. Sharda M, Khundrakpam BS, Evans AC, Singh NC. Disruption of structural covariance networks for language in autism is modulated by verbal ability. Brain Struct Funct. 2016;221: 1017–1032. doi:10.1007/s00429-014-0953-z

67. van Rooij D, Anagnostou E, Arango C, Auzias G, Behrmann M, Busatto GF, et al. Cortical and Subcortical Brain Morphometry Differences Between Patients With Autism Spectrum Disorder and Healthy Individuals Across the Lifespan: Results From the ENIGMA ASD Working Group. Am J Psychiatry. 2018;175: 359–369. doi:10.1176/appi.ajp.2017.17010100

68. Lombardo MV, Pierce K, Eyler L, Barnes CC, Ahrens-Barbeau C, Solso S, et al. Different functional neural substrates for good and poor language outcome in autism. Neuron. 2015;86: 567–577. doi:10.1016/j.neuron.2015.03.023

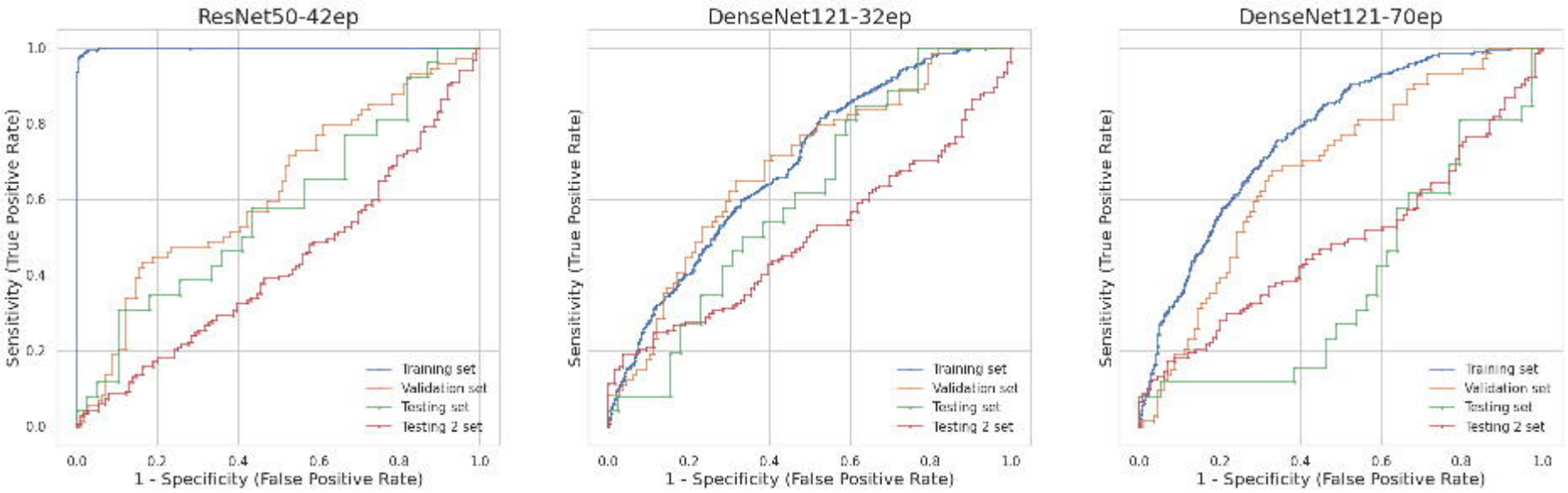
69. Szatmari P, Georgiades S, Duku E, Bennett TA, Bryson S, Fombonne E, et al. Developmental trajectories of symptom severity and adaptive functioning in an inception cohort of preschool children with autism spectrum disorder. JAMA Psychiatry. 2015;72: 276–283. doi:10.1001/jamapsychiatry.2014.2463

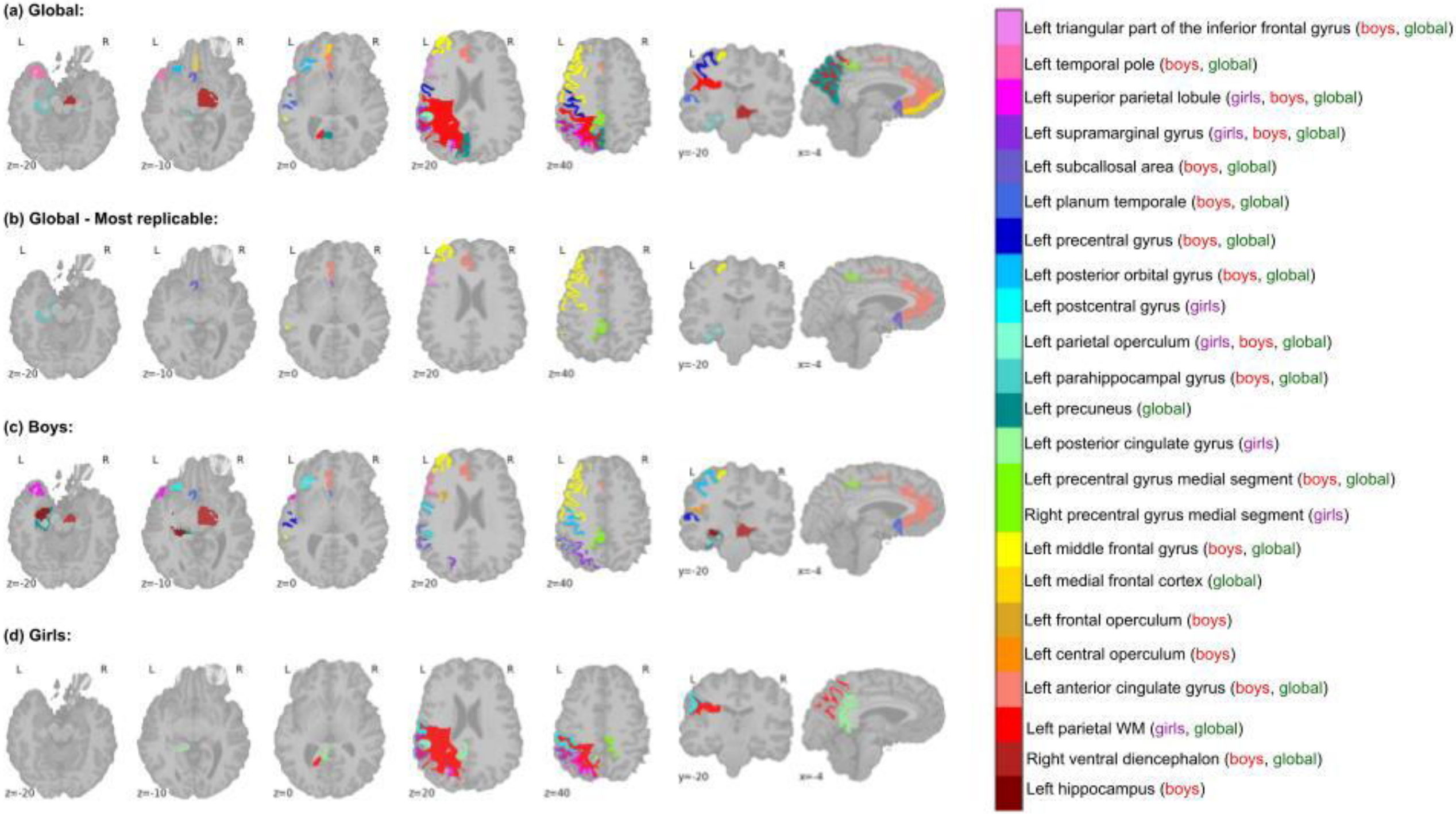
70. Tager-Flusberg H, Kasari C. Minimally verbal school-aged children with autism spectrum disorder: the neglected end of the spectrum. Autism Res. 2013;6: 468–478. doi:10.1002/aur.1329

71. Andrews DS, Avino TA, Gudbrandsen M, Daly E, Marquand A, Murphy CM, et al. In Vivo Evidence of Reduced Integrity of the Gray–White Matter Boundary in Autism Spectrum Disorder. Cereb Cortex. 2017;27: 877–887. doi:10.1093/cercor/bhw404

72. Lam P, Zhu AH, Gari IB, Jahanshad N, Thompson PM. 3D Grid-Attention Networks for Interpretable Age and Alzheimer's Disease Prediction from Structural MRI. arXiv:201109115 [eess, q-bio]. 2020 [cited 5 Jan 2021]. Available: http://arxiv.org/abs/2011.09115

73. Yao Y, Rosasco L, Caponnetto A. On Early Stopping in Gradient Descent Learning. Constr Approx. 2007;26: 289–315. doi:10.1007/s00365-006-0663-2


ResNet50-42ep
DenseNet121-32ep
DenseNet121-70ep
Sensitivity (True Positive Rate)
1 - Specificity (False Positive Rate)
0.0
0.2
0.4
0.6
0.8
1.0
Training set
Validation set
Testing set
Testing 2 set


(a) Global:
(b) Global - Most replicable:
(c) Boys:
(d) Girls:
L
R
z=-20
z=-10
z=0
z=20
z=40
y=-20
x=-4
Left triangular part of the inferior frontal gyrus (boys, global)
Left temporal pole (boys, global)
Left superior parietal lobule (girls, boys, global)
Left supramarginal gyrus (girls, boys, global)
Left subcallosal area (boys, global)
Left planum temporale (boys, global)
Left precentral gyrus (boys, global)
Left posterior orbital gyrus (boys, global)
Left postcentral gyrus (girls)
Left parietal operculum (girls, boys, global)
Left parahippocampal gyrus (boys, global)
Left precuneus (global)
Left posterior cingulate gyrus (girls)
Left precentral gyrus medial segment (boys, global)
Right precentral gyrus medial segment (girls)
Left middle frontal gyrus (boys, global)
Left medial frontal cortex (global)
Left frontal operculum (boys)
Left central operculum (boys)
Left anterior cingulate gyrus (boys, global)
Left parietal WM (girls, global)
Right ventral diencephalon (boys, global)
Left hippocampus (boys)