Ensemble learning with 3D convolutional neural networks for connectome-based prediction

4 authors, including:

Meenakshi Khosla
Cornell University
14 PUBLICATIONS   **31** CITATIONS

SEE PROFILE

Keith W Jamison
Weill Cornell Medical College
24 PUBLICATIONS   **189** CITATIONS

SEE PROFILE

Amy Kuceyeski
Weill Cornell Medical College
68 PUBLICATIONS   **748** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Geriatric Depression View project

Multiple Sclerosis View project

Ensemble learning with 3D convolutional neural networks for functional connectome-based prediction

Meenakshi Khosla[a], Keith Jamison[b], Amy Kuceyeski[b,c], Mert R. Sabuncu[a,d]

[a]*School of Electrical and Computer Engineering, Cornell University*
[b]*Radiology, Weill Cornell Medical College*
[c]*Brain and Mind Research Institute, Weill Cornell Medical College*
[d]*Nancy E. & Peter C. Meinig School of Biomedical Engineering, Cornell University*

Abstract

The specificty and sensitivity of resting state functional MRI (rs-fMRI) measurements depend on pre-processing choices, such as the parcellation scheme used to define regions of interest (ROIs). In this study, we critically evaluate the effect of brain parcellations on machine learning models applied to rs-fMRI data. Our experiments reveal a remarkable trend: On average, models with stochastic parcellations consistently perform as well as models with widely used atlases at the same spatial scale. We thus propose an ensemble learning strategy to combine the predictions from models trained on connectivity data extracted using different (e.g., stochastic) parcellations. We further present an implementation of our ensemble learning strategy with a novel 3D Convolutional Neural Network (CNN) approach. The proposed CNN approach takes advantage of the full-resolution 3D spatial structure of rs-fMRI data and fits non-linear predictive models. Our ensemble CNN framework overcomes the limitations of traditional machine learning models for connectomes that often rely on region-based summary statistics and/or linear models. We showcase our approach on a classification (autism patients versus healthy controls) and a regression problem (prediction of subject's age), and report promising results.

Keywords: Functional connectivity, fMRI, Convolutional Neural Networks, Autism Spectrum Disorder, ABIDE

1. Introduction

Functional connectivity, as often captured by correlations in resting state functional MRI (rs-fMRI) data, has produced novel insights linking differences in brain organization to individual or group-level characteristics. Recently, machine learning models are being increasingly applied to study and exploit individual variation in functional connectivity data [1, 2, 3]. These models often employ hand-engineered features, such as pairwise correlations between regions of interest (ROIs) and network topological measures of clustering, modularity, small-worldness, integration, or segregation [4, 5, 6]. The ROIs are usually computed based on a pre-defined atlas or a parcellation scheme. The choice of the ROIs can have a significant impact on downstream analyses [7, 8, 9].

Brain ROIs can be defined based on macro-anatomical features, cytoarchitecture, functional activations, and/or connectivity patterns [10, 11, 12, 13]. A common approach is to derive the ROIs either based on input from experts and/or using a data-driven strategy on a small number of subjects. Expert-defined ROIs are challenging to standardize across studies [14] and often rely on arbitrary decisions. Data-driven ROIs, on the other hand, can be biased by the selection of the subjects, especially for regions that exhibit large variability across the population. Popular data-driven techniques include clustering, dictionary learning and Independent Component Analysis (ICA) [15, 16, 9]. Such methods can be sensitive to confounds such as motion, while initialization, optimization, and other algorithmic choices can also significantly influence the results [17]. A parcellation scheme not only defines the boundaries of ROIs, but also restricts the analysis to a certain spatial scale.

Given the arbitrary nature of a chosen parcellation scheme and its impact on predictive models, we hypothesized that machine learning models can benefit markedly from an ensemble strategy that integrates across different scales and ROI definitions. Figure 1 shows a general schematic of our proposed framework. In this work, we conducted a thorough empirical evaluation of different choices for brain parcellations.

Another important factor in connectome-based machine learning pertains to the choice of the classification algorithm. A large body of related work in the literature has focused on simple linear predictive models using vectorized connectivity data. A relatively recent trend is to represent connectome data as a graph (with subjects as nodes in the graph) and harness
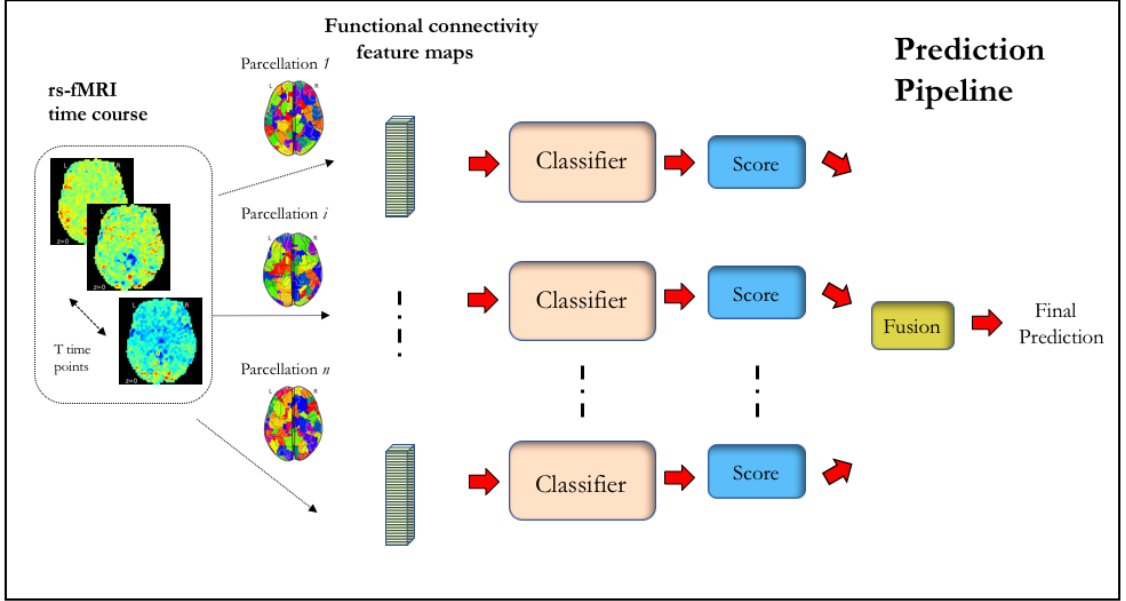
Figure 1: A general illustration of the proposed approach

graph kernels or specialized neural network architectures to build predictive models. Ktena et al. [18] applied spectral graph convolutions in a distance-metric learning framework to train a k-nearest neighbor classifier on connectivity data. In a similar vein, Kawahara et al. [19] proposed the BrainNetCNN architecture that extends convolutional neural networks (CNNs) to handle graph-structured data. While CNNs are motivated via the translation-invariance property of image-based classification problems and thus have achieved tremendous success, the neuroscientific basis of the invariance property exploited by BrainNetCNN remains elusive. Furthermore, this approach works directly with an adjacency matrix derived from the connectome data, while disregarding spatial information. As we discuss below, we propose an alternative representation of connectivity data, which allows us to leverage modern deep learning architectures, like CNNs, to build a prediction model that exploits spatial information.

In this work, we consider two applications: discrimination of autism patients and healthy controls; and regression of age. The first problem is a particularly challenging one. Several previous studies have reported altered functional connectivity patterns in Autism Spectrum Disorder (ASD) pa-

tients [20, 21, 22, 23]. While studies using small samples have reported classification accuracies over 75% [24], application of similar models on large heterogeneous datasets, such as ABIDE [25], have shown more modest performance levels over a wide range of connectome preprocessing schemes (accuracies that range 60-67%) [26].

Our main contributions in this paper are:

- An extensive evaluation of the influence of brain parcellations on functional connectome-based machine learning models

- An ensemble learning strategy for combining predictions from multiple classifiers corresponding to different brain parcellations

- An easy-to-implement 3D CNN framework for connectome-based classification

- A technique to visualize and compare trained CNN models using saliency maps that reflect the importance of individual voxels for prediction

2. Materials and Methods

2.1. Dataset

Autism Brain Imaging Data Exchange (ABIDE) is a multi-site consortium aggregating and openly sharing anatomical, functional MRI and phenotypic datasets of individuals diagnosed with ASD, as well as healthy controls (HC) [25]. The first phase of ABIDE (ABIDE-I) collected data from 1,112 individuals, comprising 539 individuals diagnosed with ASD and 573 typical controls across 17 sites. The second phase (ABIDE-II) aggregated 1,114 additional datasets, comprising 521 individuals with ASD and 593 healthy controls across 19 sites.

2.2. Preprocessing of fMRI Data

The Preprocessed Connectomes Project (PCP) released preprocessed versions of ABIDE-I using several pipelines [27]. We used the data processed through the Configurable Pipeline for the Analysis of Connectomes (CPAC). This pipeline performs motion correction, global mean intensity normalization and standardization of functional data to MNI space (3x3x3 mm resolution) before the extraction of ROI time series. Among the different strategies

4

Dataset	Prediction	Sample Size	Median Age (Range) in yrs
ABIDE-I	Age	387	13.8 (6.5-29.1)
ABIDE-I	ASD/HC	379/395	13.9 (6.5-56.2)
ABIDE-II	Age	213	10.6 (5.8-18.8)
ABIDE-II	ASD/HC	163/230	11.0 (5.2-38.9)

Table 1: Composition of Cohorts

in the release, our analysis used data de-noised by regression of nuisance signals including motion parameters, CompCor WM+CSF components, and global signal, followed by band-pass filtering (0.01-0.1Hz).

We preprocessed the ABIDE-II dataset following the same sequence of steps listed for ABIDE-I in CPAC (using the version v1.0.2a). Since manual quality control (QC) was not yet available for ABIDE-II, we performed an automatic QC by selecting those subjects that retained at least 100 frames or 4 minutes of fMRI scans after motion scrubbing [28]. Motion scrubbing was performed based on Framewise Displacement (FD), discarding one volume before and two volumes after the frame with FD exceeding 0.5mm [29].

2.3. Cohort selection

In our experiments, we used ABIDE-I subject data that passed manual QC by all the functional raters. This yielded a final sample size of 774 ABIDE-I subjects, comprising 379 subjects with ASD and 395 typical controls. As an independent test dataset, we employed ABIDE-II subjects from sites that participated in ABIDE-I and used the same MRI sequence parameters for data collection. After automatic QC, we ended up with a final ABIDE-II sample size of 163 individuals with ASD and 230 healthy controls. For age prediction, we only considered healthy controls. Furthermore, subjects whose age were more than 3.5 standard deviations away from the median were excluded from the task of age prediction. Table 1 summarizes the dataset characteristics for the two prediction tasks considered in this study.

2.4. Extracting ROI time series from atlases

In our experiments, we considered all atlases that were used for ROI time series extraction in PCP. These include the following seven atlases: Talaraich and Tournoux (TT, R=97), Harvard-Oxford (HO, R=111), Automated Anatomical Labelling (AAL, R=116), Eickhoff-Zilles (EZ, R=116),

5

Dosenbach 160 (DOS160, R=161), Craddock 200 (CC200, R=200), and Craddock 400 (CC400, R=392), where R is the number of ROIs [30, 31, 32, 33, 34, 35, 36, 37, 38].

For our 3D CNN model, described below, the parcellated regions were used as target ROIs to derive the input connectivity features at the voxel level. For the non-CNN benchmark models, also described below, each atlas was used to define a corresponding connectivity matrix which was fed as input to each model after collapsing into a vector. We report results for ensemble learning strategies as well, where we combined the predictions of models corresponding to individual atlases.

2.5. Creating stochastic parcellations

Stochastic parcellations were created by Poisson Disk Sampling using the method described in [39]. Given a number of ROIs, this approach divides the gray matter voxels (as defined by a given mask) into roughly equal-sized parcels while ensuring that the parcels do not cross hemisphere boundaries. Stochasticity is introduced in the ROI center locations, and all the remaining voxels are assigned to the closest region center. These centers are kept a minimum distance apart based on the desired number of regions in the parcellation. All parcellations were created in the MNI152 template at a 3mm resolution, same as the resolution of the preprocessed functional data. For creating these parcellations, we relied on a whole brain gray matter mask including sub-cortical structures. To create the mask, we took the union of the gray matter tissue prior provided in the standard MNI152 template and the cortical mantle mask used in [16].

2.6. 3D Convolutional Neural Network Approach

Here, we present our novel strategy to adopt a 3D CNN architecture for use with connectomic data. The input to the CNN is formed by concatenating voxel-level maps of "connectivity fingerprints", which are represented as a multi-channel 3D volume. Each channel is a connectivity feature, such as the Pearson correlation between each voxel's time series and the average signal within a target ROI. In our implementation, we use both atlas-based and stochastic brain parcellation schemes to define target ROIs. The total number of input channels thus represents the number of ROIs used for creating voxel-level fingerprints. For each parcellation scheme (atlas-based or stochastic), we trained a separate model.

In our experiments, we employed a simple CNN architecture, illustrated in Fig. 2. Our architecture has several convolutional layers, interspersed with max-pooling based down-sampling layers, followed by a couple of densely connected layers. The models were trained with a mini-batch size of 64, until convergence of validation loss. For classification, we used binary cross-entropy, whereas for regression we adopted absolute difference as the loss function.

The neural network weights were then optimized via stochastic gradient descent (SGD). The learning rate and momentum for SGD were set to 0.001 and 0.9 respectively. The same architecture and settings were used for all atlases and stochastic parcellations. We note that each atlas is defined on a unique gray matter mask. To ensure that all prediction models (benchmark and proposed) relied on information from the same voxels, the atlas-specific gray matter mask was applied to the voxel-level connectivity fingerprint data before feeding into the proposed convolutional architecture. For stochastic parcellations, the custom gray matter mask as described above was used for masking the fingerprints.
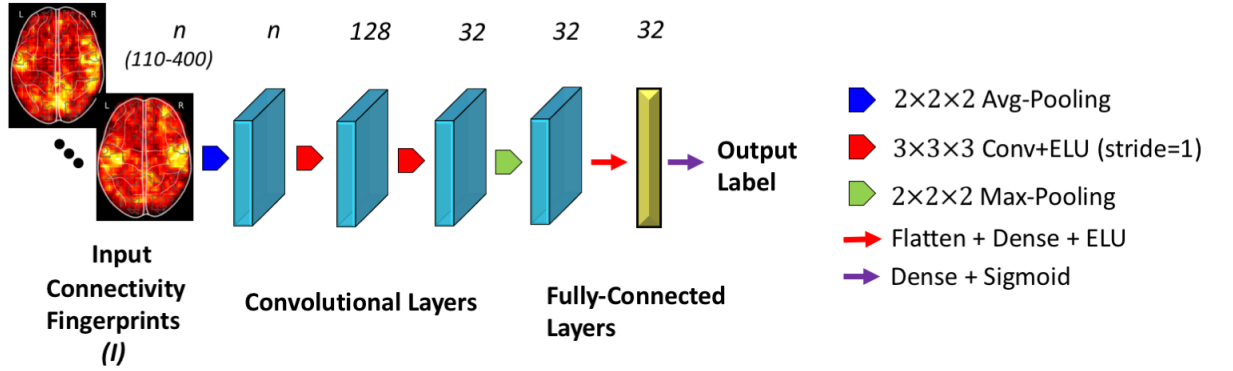

Figure 2: Proposed CNN approach. All operations are in 3D volume. 2D correlation maps are shown for illustration only. For the age prediction task, an additional Max-Pooling operation followed the first convolutional layer.

2.7. Benchmark Methods

In our experiments, we implemented following benchmark methods.

2.7.1. Ridge Regression

A linear regression model was trained with squared loss and α times the squared norm of the weight vector (See Appendix). For classification, the ground truth labels were encoded as ± 1 for the two output categories. We tested 10 linearly spaced values for the hyper-parameter α in the range [0.1,10] and report the highest cross-validation accuracy. Thus this baseline result reflects an *optimistic* estimate of performance.

2.7.2. Support Vector Machine

We implemented a standard SVM as a benchmark (See Appendix). We found that a radial basis function (RBF) kernel performed better than a linear model. Thus we report results for the RBF-kernel SVM. The two hyper-parameters (RBF kernel width γ and and misclassification cost weight C) were fine-tuned by maximizing cross-validation accuracy via a grid search. Therefore, as with the ridge classifier, this should be considered as an upper bound on generalization performance. For regression, we implemented the standard SVR scheme with an ϵ- insensitive loss function, optimizing for the ϵ-tube and penalty parameter of the error term via grid search.

2.7.3. Fully Connected Architecture

The fully-connected neural network (FCN) architecture takes as input functional connectivity estimates between pairs of ROIs, which is vectorized and processed by a feed-forward network. We implemented following architecture, which performed best on ABIDE-I cross-validation: 4 fully connected hidden layers, with 800, 500, 100 and 20 numbers of features and each linear layer followed by an elementwise Exponential Linear Unit (ELU) activation. Dropout regularization parameter was set to 0.2 and applied to each layer during training. For classification, the output node was a sigmoid, and cross-entropy loss was used. For age prediction, the sigmoidal output was replaced with a linear activation and absolute difference was used as the loss function. The models were trained with a mini-batch size of 64, until convergence of validation loss.

SGD was used as the optimizer with learning rate and momentum set to 0.01 and 0.9 respectively for classification. For age prediction, a smaller learning rate of 0.002 was used.

2.7.4. BrainNet Convolutional Neural Networks

BrainNet CNN, originally proposed in [19], utilizes specialized kernels to handle connectomic data. Their work described novel edge-to-edge, edge-to-node and node-to-graph convolutional layers that can potentially capture topological relationships between network edges. For BrainNet CNN, we implemented the following architecture that worked best on ABIDE-I cross-validation: 1 edge-to-node layer with 256 filters, followed by a node-to-graph layer with 128 output nodes and finally a dense layer with single output. A leaky ReLU non-linearity with alpha equal to 0.33 was applied to the output of each layer except the last layer. The activation of the last layer was set to linear and sigmoid for the regression and classification tasks, respectively. Dropout regularization with rate 0.2 was used for the edge-to-node layer. Similar to [19], Euclidean loss was minimized for age regression, whereas cross-entropy loss was used to optimize the classification models. The models were trained for 1000 iterations using SGD with momentum equal to 0.9. The learning rate was set to 0.0005 for age prediction and 0.008 for ASD/Healthy classification. The training curves were monitored for atlases to ensure convergence.

2.8. Ensemble Learning

In our experiments, we explored two ensemble learning strategies. The first one is what we call multi-atlas ensemble (or MA-Ensemble). MA-Ensemble averages the predictions of the models of a specific method (e.g., BrainNet CNN) computed using each one of the seven atlases. For classification, the final prediction is computed as the majority vote of the individual binary class predictions. For regression, the ensemble prediction is simply the mean. The second ensemble strategy (SP-Ensemble) averages across the models of a specific method computed using stochastic parcellations. In our experiments, unless stated otherwise, we used 30 stochastic parcellations at each of the following four spatial scales: 110, 160, 200 and 400 ROIs. These scales were chosen in accordance with existing atlases. Thus the SP-Ensemble's prediction was computed based on fusing 120 (30×4 scales) models. We also implemented single-scale SP-Ensemble models, which averaged over the 30 parcellations at the same spatial scale.

2.9. Visualizing the CNN model

In order to understand the connectivity features captured by the CNN model, we employed the saliency map approach of [40]. This visualization

Parcellation	Ridge	SVM	FCN	BrainNet	3D-CNN
HO	66.7/63.3	69.4/**68.7**	69.4/67.7	67.8/66.1	**70.5**/67.7
CC200	69.7/67.4	69.1/70.7	70.5/*71.5*	68.6/70.2	**71.2/*72.8***
EZ	66.4/63.3	69.0/66.1	68.6/63.8	66.0/64.4	**69.3/66.4**
TT	64.4/66.1	68.6/67.4	67.1/65.9	66.0/67.4	**69.4/70.0**
CC400	70.2/69.4	69.4/68.2	71.0/69.9	71.3/***71.5***	**71.7**/70.5
AAL	65.4/63.3	69.1/65.9	66.7/65.4	66.5/64.6	**71.4/69.5**
DOS160	66.2/66.7	68.4/63.6	67.2/66.1	67.0/64.6	**68.6/67.0**
MA-Ensemble	69.8/66.7	70.5/70.0	71.5/69.9	69.7/70.7	**73.3/71.7**
SP-Ensemble	*70.7/71.7*	*71.0*/71.2	*72.0*/71.2	*71.5*/70.5	***73.5/72.3***

Table 2: Classification accuracy for ASD vs. Control: 10-fold cross-validation on ABIDE-I/independent test on ABIDE-II accuracy of baseline models and proposed CNN approach. For each row, best results are **bolded**. For each column, best results are *italicized*.

technique computes the gradient of the output prediction with respect to the input image voxel values, i.e., the 3D volume, using a single backward pass through the trained neural network. We then computed voxel-level saliency as the maximum absolute gradient value across all input channels corresponding to different target ROIs. More formally, consider an input image $\mathbf{I}$, representing the connectivity fingerprints of v voxels with R ROI signals. The saliency weights $w \, \epsilon \mathbb{R}^{v \times R}$ are computed by taking the absolute value of the gradient of neural network output O with respect to the input image, i.e., $w = |\frac{\partial O}{\partial I}|$. In order to obtain the saliency at the voxel level $S \, \epsilon \, \mathbb{R}^v$, we take the maximum across all the ROIs, i.e., $S_i = \max_{1 \leq j \leq R} w_{ij}$. Finally, to visualize an ensemble model, we averaged the individual saliency maps that made up the ensemble.

3. Results

3.1. Experiments

In our experiments, we considered two tasks: i) binary classification of autism vs healthy, and ii) age prediction. For each task, we implemented two evaluation schemes. First, we conducted 10-fold cross-validation on the ABIDE-I dataset, so that we could present results that were consistent with previously reported classification results such as [1, 26]. Second, we trained each model on the entire ABIDE-I dataset and computed test performance on the independent ABIDE-II set. We report classification accuracy and

Parcellation	Ridge	SVM	FCN	BrainNet	3D-CNN
HO	2.60/2.39	2.90/2.40	2.62/2.28	2.64/2.31	**2.54/2.25**
CC200	2.54/2.22	2.90/2.26	2.64/2.22	2.59/**2.17**	**2.52**/2.20
EZ	2.68/2.35	*2.87/2.22*	2.75/2.25	2.69/2.37	**2.55/2.11**
TT	2.67/2.44	2.95/2.33	2.65/2.44	2.68/2.40	**2.60/2.08**
CC400	2.58/2.28	2.95/2.40	**2.56**/2.19	2.59/**2.10**	2.57/2.11
AAL	2.72/2.26	*2.87*/2.25	2.83/2.39	2.76/2.19	**2.48/2.07**
DOS160	2.76/2.82	3.16/2.90	2.78/2.75	2.74/2.75	**2.70/2.61**
MA-Ensemble	*2.47*/2.22	2.90/2.34	*2.42*/2.22	2.45/2.14	2.43/**2.08**
SP-Ensemble	2.50/*2.18*	2.89/2.24	*2.42/2.00*	*2.43/2.06*	***2.41/1.94***

Table 3: Mean absolute error (MAE in years) for age prediction: 10-fold cross-validation on ABIDE-I healthy subjects/independent test on ABIDE-II for benchmark models and proposed CNN approach. For each row, best results are **bolded**. For each column, best results are *italicized*.

the receiver operating curves (ROC), along with corresponding area under the curves (AUC) for each of these scenarios under various combinations of parcellation schemes and prediction algorithms. For age prediction, we report the Mean Absolute Error (MAE).

3.2. Evaluation of Prediction Performance

Table 2 shows the cross-validation and independent test performance for different models on the classification problem. Clearly, the proposed 3D CNN approach shows superior prediction performance compared to the benchmark methods, including fully-connected deep neural network (FCN) and BrainNet CNN. In particular, the 3D CNN approach yields the best ABIDE-I cross-validation accuracy for all parcellation schemes, including the ensembles. Similarly, the SP-Ensemble achieves the best ABIDE-I cross-validation for all algorithms, including the 3D CNN. The ABIDE-II results are in general compatible with the cross-validation results, where the 3D CNN and SP-Ensemble techniques mostly outperform the competition. Figure 3 shows the Receiver Operating Characteristic (ROC) curves for SP-Ensemble models for the different algorithms on the independent ABIDE-II test dataset. We observe that the 3D CNN achieves the largest AUC of 76.86%.

Table 3 lists cross-validation and independent test results for the age prediction task. The 3D CNN approach consistently shows superior performance, yielding the best results for all but two parcellation schemes. Sim-

Receiver Operating Characteristic Curves

Legend:
- SP-Ensemble CNN (area = 0.7686)
- SP-Ensemble FCN (area = 0.7636)
- SP-Ensemble BrainNet CNN(area = 0.7557)
- SP-Ensemble Ridge (area = 0.7412)
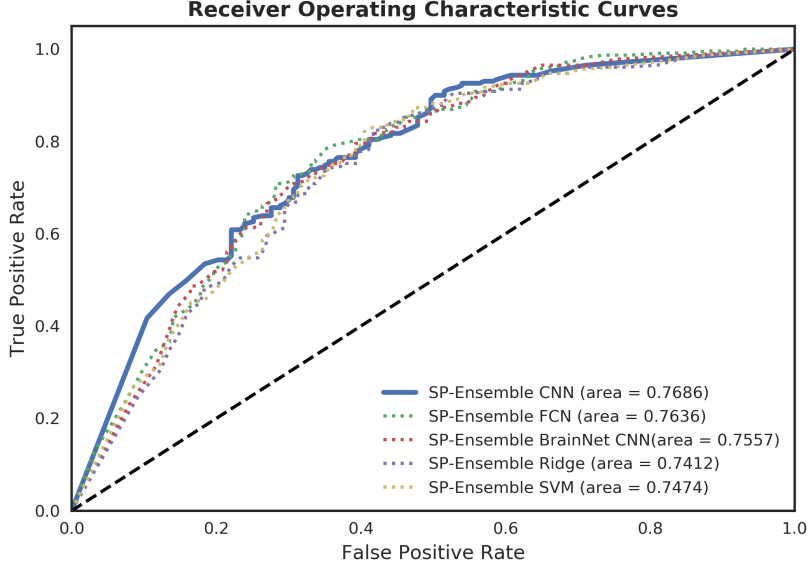- SP-Ensemble SVM (area = 0.7474)

Figure 3: ASD-HC Classification: Receiver Operating Curves for independent validation on ABIDE-2

ilar to the classification scenario, SP-Ensemble also yields the best cross-validation and independent test performance values for the majority of the algorithms, including 3D CNN. Overall, the best accuracy is achieved by SP-Ensemble 3D CNN, which yields a mean absolute error of **2.41** years on ABIDE-I cross-validation and **1.94** years on the independent ABIDE-II dataset.

3.3. Comparison of stochastic parcellations and atlases

Here, our objective is to conduct a detailed investigation of how the choice of ROIs affects prediction performance for different machine learning (ML) algorithms. For each ML algorithm and each parcellation we have a model trained on the ABIDE-I data, which we then used on the independent ABIDE-II data to quantify prediction accuracy. Figure 4 shows the distribution of accuracy values (estimated with a kernel density model) obtained using stochastic parcellations , while also illustrating the results for each of the atlases and the scale-specific SP-ensembles. The scale-specific SP-Ensemble strategy, as the name implies, averaged the models corresponding to the 30
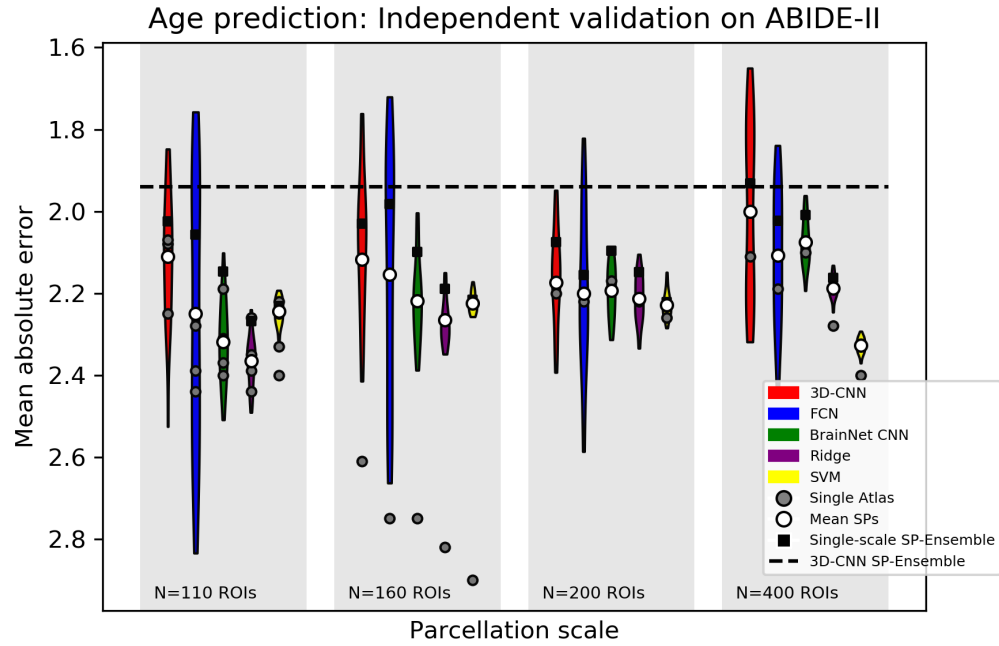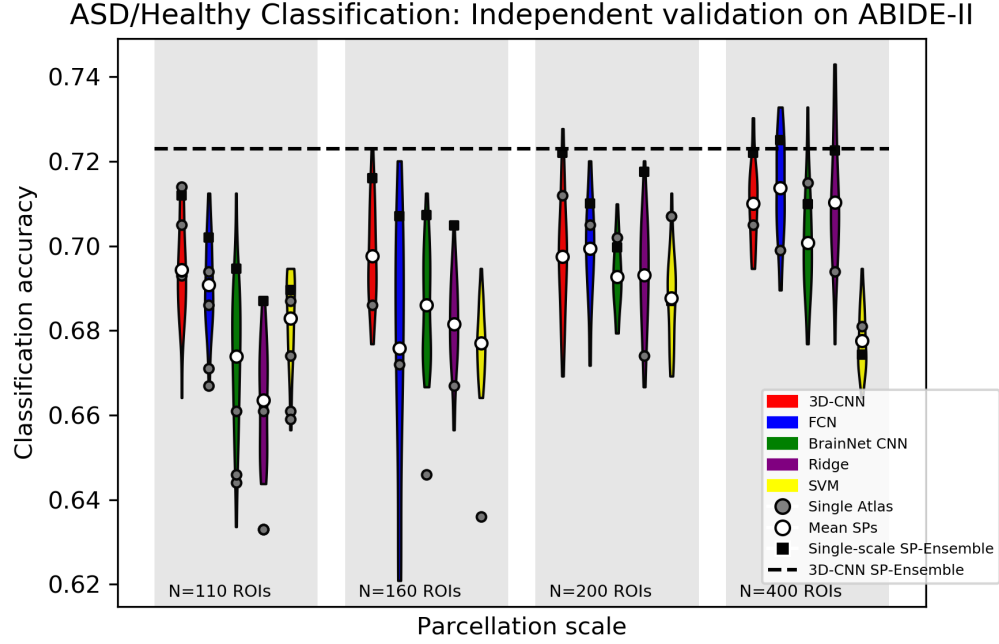
Figure 4: Violin plots showing the spread of prediction accuracies/errors for stochastic parcellations at multiple network scales for different classification models. Mean accuracy/error of individual violins is denoted by 'Mean SPs'. Performance of individual atlases is compared with SPs with the closest # of ROIs and is denoted as 'Single Atlas'. Results are computed by training models on entire ABIDE-1 cohort and testing on the independent ABIDE-2 cohort.
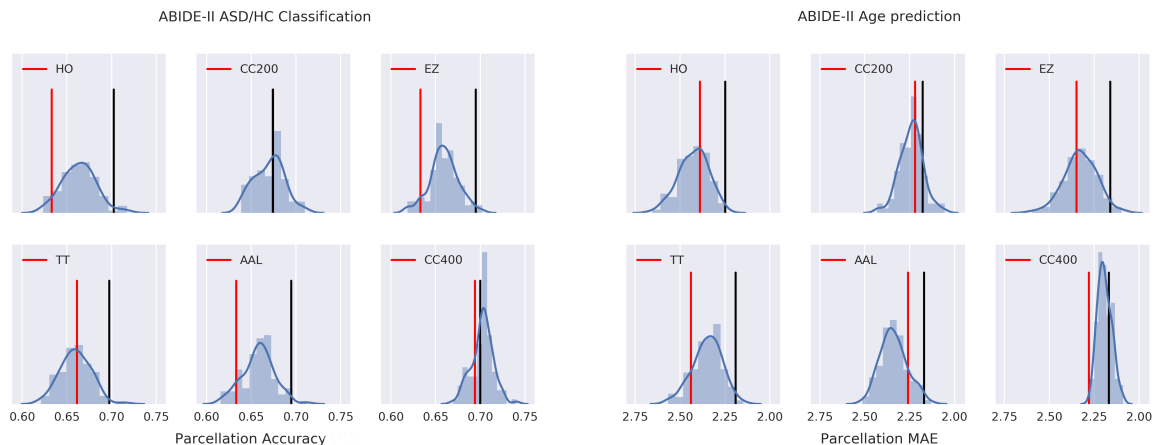
ABIDE-II ASD/HC Classification

ABIDE-II Age prediction

Figure 5: Distribution of Ridge models' performance for stochastic parcellations created using the same gray-matter mask as the corresponding atlas. Red denotes the atlas model's accuracy and black indicates the SP-Ensemble accuracy.

stochastic parcellations in each scale. We observe that the atlas-based models performed no better than typical stochastic parcellation models, independent of scale and algorithm.

This result presents a revelatory perspective: perhaps we do not need anatomically or functionally derived brain parcellations for machine learning. Furthermore, the proposed SP-Ensemble CNN strategy yielded accuracy results that were about as good as the best scale-specific SP-Ensemble model. Finally, the ensemble models were almost always better than the atlas-based models and they compared favorably against the individual stochastic parcellation models. The same observations can be made for ABIDE-I cross-validation (see Supplementary Figure S.1).

In above analysis, one potential confound was the different gray matter masks of atlases and stochastic parcellations (SPs). In order to account for this confound, we conducted following analysis. For each of the atlases, we generated 100 SPs using the same gray matter mask as the atlas. We excluded DOS160 because it does not rely on a well-defined gray matter mask and places discontiguous 4.5 mm spherical regions over fixed coordinates in the brain (sampling only 5% of brain voxels). We then trained on each of these SPs using the same hyper-parameters that were found to be optimal for the corresponding atlas. Here, we show the results for ridge regression (the model that was fastest to train), but we obtained similar results for all
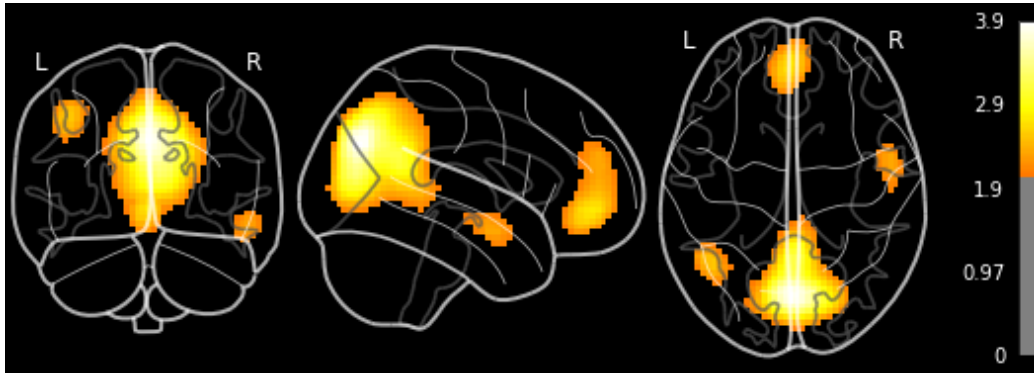
(a) ASD/Healthy Classification
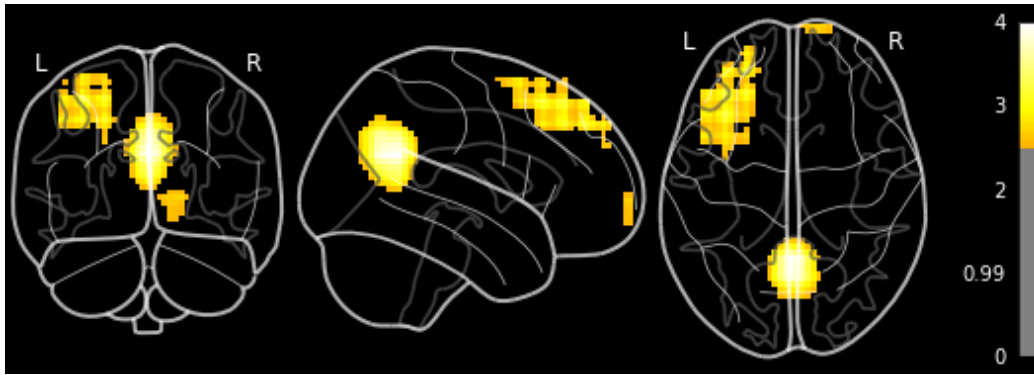
(b) Age prediction

Figure 6: Mean saliency maps of trained 3D-CNN models for SP-Ensemble

other algorithms as well. As can be seen from Figure 5, for most atlases and corresponding gray matter masks, the model trained on the atlas ROIs performed no better than an average SP model. Furthermore, and importantly, the SP-Ensemble (computed by averaging across SPs on the atlas-specific mask) yielded better performance than the atlas models for all atlases.

3.4. Visualization

An important goal of machine-learning tools in neuroimaging is to generate novel insights linking imaging biomarkers with disease or phenotypic traits. Visualization techniques for CNNs can help reveal important features used by the model for discriminating between output classes. Figure 6 shows the saliency maps computed for the SP-Ensemble CNN ASD classification and age prediction models. As can be seen from these maps, the precuneus,

often considered a core node of the default mode network [41], seems to play a significant role for both prediction problems. However, there are also salient regions that are unique to each problem. For example, the anterior cingulate/ventromedial prefrontal cotex, a region that has been linked to autism [42], was distinctly highlighted for the ASD classification problem. The left parietal cortex was also emphasized for ASD prediction, which is consistent with the laterilized activation observed in this region in Autism patients [43]. On the other hand, for age prediction, the left dorsolateral prefrontal cortex (dlPFC) is a uniquely salient region. The dlPFC is associated with executive functions, such as working memory and abstract reasoning. For working memory, dlPFC's function seems to be age-associated and more lateralized in younger adults [44].

4. Discussion

In this study, we presented a detailed empirical analysis of how the choice of ROIs can impact the performance of machine learning models trained on functional connectomes. We considered several machine learning algorithms, together with a range of spatial scales and parcellation schemes, including the popular atlas-based techniques and a stochastic approach. Our analysis suggests that using a single atlas for summarizing the connectome data is often sub-optimal for training machine learning models, and significantly more accurate predictions can be achieved with an ensemble approach that averages across models trained with different parcellation schemes. Furthermore, we demonstrated that averaging across stochastic parcellations can achieve very high accuracy values, often surpassing atlas-based models.

Another main contribution of this study is a novel approach to employ a 3D CNN architecture on functional connectivity data. Convolutional neural networks achieve state-of-the-art performance on many image-based prediction tasks, as they take advantage of the full spatial resolution of the data and the translation invariance property of the problem. Our proposed approach treats voxel-level connectivity fingerprints as input channels to a conventional 3D CNN framework. This strategy contrasts with widely used machine learning models that work with functional connectivity data, while ignoring the spatial structure. Our results demonstrate that when tailored for connectomes, CNNs offer a promising opportunity to probe brain networks in disease.

16

Machine learning practitioners have to make a number of preprocessing choices in extracting connectomic features to analyze. While there is no one-size-fits-all solution across different tasks, in the context of machine learning models of functional connectivity, we present some interesting empirical observations below.

4.1. Ensemble learning

The motivation behind using multiple stochastic parcellations for prediction is grounded in the concept of ensemble learning. The core idea is to integrate out a latent variable (i.e., parcels or ROI definitions) from the learning problem [45]. This approach also makes the predictions more robust to the precise parcellation scheme. As shown above, the performance of atlas-based models can vary significantly ($\sim$5-10% for parcellations at the same scale). In such a scenario, ensemble learning over multiple stochastic parcellations can be a robust strategy that yields reliable predictions.

4.2. Network granularity

We explored the impact of network granularity on prediction performance of machine learning algorithms for connectomes. Our analysis suggests that better prediction performance can be expected with parcellations at higher granularity. However, no significant differences were observed beyond 400 ROIs. Our evaluations contradict with a previously reported result that a coarser network scale ($\sim$ 100-150 ROIs) is more suitable for autism classification [46]. In their paper, these conclusions were drawn by comparing the performances achieved with a few atlases. However, inferring trends from a small number of atlases can be misleading, since factors like the boundary definitions of structures (cortical/subcortical) or the particular gray matter mask used, will effect results. Stochastic parcellations can control for these confounds and depict unbiased trends across network scales.

4.3. Number of gray matter voxels

Our empirical study suggests that there is no direct correlation between the number of voxels in the gray matter mask and a model's prediction performance. However, we do observe that the choice of gray matter mask can impact results. For example, the DOS160 atlas with as few as $\sim$ 3,039 voxels shows performance no worse than other atlases at the same resolution (HO, EZ, TT and AAL) with $\sim$ 20x more voxels.

Figure 7: Motion correlations
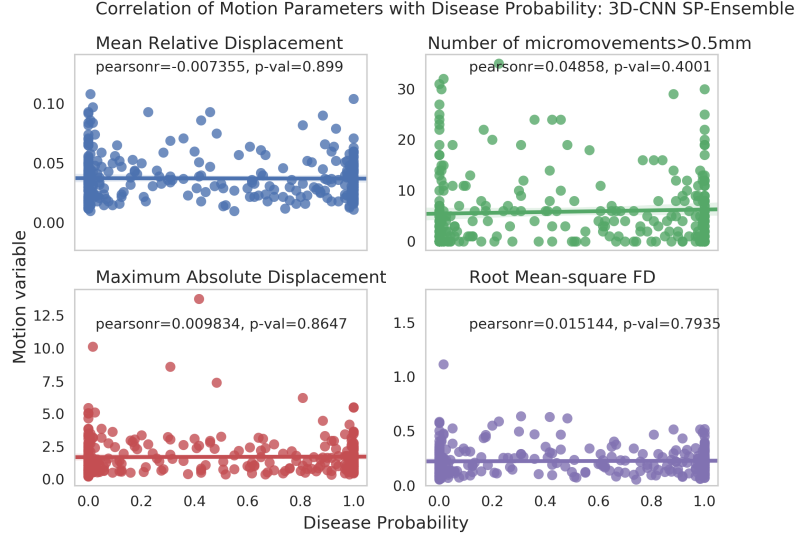
4.4. Visualization

Saliency maps provide a valuable visualization strategy to probe deep neural network models. We visualized the saliency maps from 3D CNN models trained on ROIs extracted using both atlases and stochastic parcellations. As shown in Figure 6 and Supplementary Figures S.2 and S.3, these maps are remarkably consistent.

These maps reveal that the precuneus, which is a hub of the default mode network and associated with ASD and age, plays an important role for both prediction problems. There were also uniquely highlighted regions, such as the anterior cingulate/ventromedial prefrontal cotex for ASD classification and the left dorsolateral prefrontal cortex (dlPFC) for age prediction.

4.5. Influence of motion

Several studies have shown differences in head motion parameters during fMRI between healthy controls and diseased populations, or between subjects from different age groups [47, 48]. This, in turn, can manifest as artifacts in the derived resting-state connectivity [49]. Although our independent test data was motion scrubbed, we performed additional analyses to rule out the confounding effect of motion in classifier decisions. We selected a cohort of

18

151 ASD subjects with motion-matched healthy controls from our independent dataset and analyzed the correlation of 4 motion parameters with classifier predictions. These include the root-mean-square framewise displacement, mean relative displacement, maximum absolute displacement and the number of micro-movements greater than 0.5mm. These summary statistics were chosen in accordance with previous reports of motion artifacts in rs-fMRI[50]. As shown in Figure 7, no significant correlations were observed between motion variables and the predictions of SP-Ensemble (model average over all atlases). In this motion-matched cohort, classification accuracy of 71.8% was obtained using 3D-CNN.

For our regression task, there was no significant correlation between a subject's age and any of these motion parameters in our cohorts.

4.6. Limitations and future work

Throughout our analysis, Pearsons correlation was chosen to measure functional connectivity strength between different brain regions. Several other correlation metrics, including tangent-based and partial correlation have been shown to yield superior classification performance in prior studies [9, 26]. While we do not expect this to affect the general conclusions and findings of our study, the choice of the correlation metric still remains an arbitrary decision in any machine learning pipeline for connectomes.

Due to the heavy computational burden required for training multiple deep learning models, we only considered one particular scheme for creating stochastic parcellations, i.e., Poisson Disk Sampling. Alternative strategies for creating random parcellations have also been proposed, for instance, through stochastic sub-division of anatomically derived ROIs into smaller parcels [51]. It is also possible to randomize several other more popular schemes for parcellating the brain, such as, using Ward's clustering on functional data from sub-samples of the population [45] or creating Geometric parcellations with different initializations [13].

While the proposed CNN approach achieves promising accuracy on autism detection and age prediction, there is room for further improvement. We have not yet conducted a comprehensive optimization of the convolutional architecture. Furthermore, there are likely more optimal choices than target ROI-based correlations that are used as input to the model. An interesting alternative would be select random gray matter vertices for connectivity profiling, as proposed in [16]. We envision an end-to-end learning strategy that can enable the optimization of these connectomic features.

Saliency maps provide an appealing visualization technique by mapping the neural network activations back to input voxel space. Several modifications to gradient-based back-propagation have been reported in literature that can potentially highlight more informative features learnt by the model [52, 53]. Further, the use of saliency maps need not be restricted to depicting group-averaged discriminative features. Unsupervised learning on saliency maps can provide novel insights into clinical subtypes of disease. It is also important to note that machine learning techniques do not unequivocally provide evidence for the salient features being directly associated with the disease or other target variables. However, when combined with detailed future investigations, they can spur clinical discoveries.

4.7. Conclusion

The results presented in our paper showcase the utility of ensemble learning for connectomes. Functional network based prediction models are impacted by several a priori choices, the most pivotal of which is the ROI definition. We demonstrate that ensembles of stochastic parcellations yield predictions that are significantly more robust and accurate compared to single atlas-based approaches. Further, our experiments highlight the potential of convolutional neural network models for connectome-based classification.

5. Acknowledgements

6. References

[1] M. Plitt, K. A. Barnes, A. Martin, Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards, Neuroimage Clin 7 (2015) 359–366.

[2] M. Mennes, N. Vega Potler, C. Kelly, A. Di Martino, F. X. Castellanos, M. P. Milham, Resting state functional connectivity correlates of inhibitory control in children with attention-deficit/hyperactivity disorder, Front Psychiatry 2 (2011) 83.

[3] G. Varoquaux, F. Baronnet, A. Kleinschmidt, P. Fillard, B. Thirion, Detection of brain functional-connectivity difference in post-stroke patients using group-level covariance modeling, Med Image Comput Comput Assist Interv 13 (2010) 200–208.

[4] C. J. Brown, G. Hamarneh, Machine learning on human connectome data from MRI, CoRR 1611.08699 (2016).

[5] M. Kaiser, A Tutorial in Connectome Analysis: Topological and Spatial Features of Brain Networks, ArXiv e-prints (2011).

[6] A. Alexander-Bloch, P. E. Vértes, R. Stidd, F. Lalonde, L. Clasen, J. L. Rapoport, J. N. Giedd, E. T. Bullmore, N. Gogtay, The anatomical distance of functional connections predicts brain network topology in health and schizophrenia., Cerebral cortex 23 1 (2013) 127–38.

[7] S. M. Smith, K. L. Miller, G. Salimi-Khorshidi, M. Webster, C. F. Beckmann, T. E. Nichols, J. D. Ramsey, M. W. Woolrich, Network modelling methods for FMRI (2011).

[8] Z. Yao, B. Hu, Y. Xie, P. Moore, J. Zheng, A review of structural and functional brain networks: small world and atlas, Brain Informatics 2 (2015) 45–52.

[9] K. Dadi, M. Rahim, A. Abraham, D. Chyzhyk, M. Milham, B. Thirion, G. Varoquaux, Benchmarking functional connectome-based predictive models for resting-state fMRI, 2018. Working paper or preprint.

[10] B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. Van Der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, et al., Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain, Neuron 33 (2002) 341–355.

[11] M. F. Glasser, T. S. Coalson, E. C. Robinson, C. D. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C. F. Beckmann, M. Jenkinson, et al., A multi-modal parcellation of human cerebral cortex, Nature 536 (2016) 171–178.

[12] S. B. Eickhoff, B. Thirion, G. Varoquaux, D. Bzdok, Connectivity-based parcellation: Critique and implications, Human brain mapping 36 (2015) 4771–4792.

[13] S. Arslan, S. I. Ktena, A. Makropoulos, E. C. Robinson, D. Rueckert, S. Parisot, Human brain mapping: A systematic comparison of parcellation methods for the human cerebral cortex, NeuroImage 170 (2018) 5 – 30. Segmenting the Brain.

[14] P. A. Yushkevich, R. S. Amaral, J. C. Augustinack, A. R. Bender, J. D. Bernstein, M. Boccardi, M. Bocchetta, A. C. Burggren, V. A. Carr, M. M. Chakravarty, et al., Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo mri: towards a harmonized segmentation protocol, Neuroimage 111 (2015) 526–541.

[15] G. Varoquaux, A. Gramfort, F. Pedregosa, V. Michel, B. Thirion, Multisubject dictionary learning to segment an atlas of brain spontaneous activity, in: Biennial International Conference on Information Processing in Medical Imaging, Springer, pp. 562–573.

[16] B. T. Thomas Yeo, F. M. Krienen, J. Sepulcre, M. R. Sabuncu, D. Lashkari, M. Hollinshead, J. L. Roffman, J. W. Smoller, L. Zllei, J. R. Polimeni, B. Fischl, H. Liu, R. L. Buckner, The organization of the human cerebral cortex estimated by intrinsic functional connectivity, Journal of Neurophysiology 106 (2011) 1125–1165. PMID: 21653723.

[17] B. Thirion, G. Varoquaux, E. Dohmatob, J.-B. Poline, Which fmri clustering gives good brain parcellations?, Frontiers in neuroscience 8 (2014) 167.

[18] S. I. Ktena, S. Parisot, E. Ferrante, M. Rajchl, M. Lee, B. Glocker, D. Rueckert, Metric learning with spectral graph convolutions on brain connectivity networks, NeuroImage 169 (2018) 431 – 442.

[19] J. Kawahara, C. Brown, S. P Miller, B. Booth, V. Chau, R. Grunau, J. Zwicker, G. Hamarneh, Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment 146 (2016).

[20] V. L. Cherkassky, R. K. Kana, T. A. Keller, M. A. Just, Functional connectivity in a baseline resting-state network in autism., Neuroreport 17 16 (2006) 1687–90.

[21] M. Assaf, K. Jagannathan, V. Calhoun, L. Miller, M. Stevens, R. Sahl, J. O'Boyle, R. Schultz, G. Pearlson, Abnormal functional connectivity

of default mode sub-networks in autism spectrum disorder patients 53 (2010) 247–56.

[22] C. S. Monk, S. J. Peltier, J. L. Wiggins, S.-J. Weng, M. Carrasco, S. Risi, C. Lord, Abnormalities of intrinsic functional connectivity in autism spectrum disorders,, NeuroImage 47 (2009) 764 – 772.

[23] A. S. Heinsfeld, et al., Identification of autism spectrum disorder using deep learning and the abide dataset, in: NeuroImage: Clinical.

[24] N. Yahata, J. Morimoto, R. Hashimoto, G. Lisi, K. Shibata, K. Y, A small number of abnormal brain connections predicts adult autism spectrum disorder, Nature Communications 7 (2016).

[25] A. Di Martino, D. O'Connor, B. Chen, K. Alaerts, J. S. Anderson, M. Assaf, J. H. Balsters, L. Baxter, A. Beggiato, S. Bernaerts, L. M. E. Blanken, S. Y. Bookheimer, B. B. Braden, L. Byrge, F. X. Castellanos, M. Dapretto, R. Delorme, D. A. Fair, I. Fishman, J. Fitzgerald, L. Gallagher, R. J. J. Keehn, D. P. Kennedy, J. E. Lainhart, B. Luna, S. H. Mostofsky, R.-A. Mller, M. B. Nebel, J. T. Nigg, K. O'Hearn, M. Solomon, R. Toro, C. J. Vaidya, N. Wenderoth, T. White, R. C. Craddock, C. Lord, B. Leventhal, M. P. Milham, Enhancing studies of the connectome in autism using the autism brain imaging data exchange ii, Scientific data 4 (2017) 170010.

[26] A. Abraham, M. P. Milham, A. D. Martino, R. C. Craddock, D. Samaras, B. Thirion, G. Varoquaux, Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example, NeuroImage 147 (2017) 736 – 745.

[27] C. Craddock, Y. Benhajali, C. Chu, F. Chouinard, A. Evans, A. Jakab, B. S. Khundrakpam, J. D. Lewis, Q. Li, M. Milham, C. Yan, P. Bellec, The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivative, Frontiers in Neuroinformatics (2013).

[28] J. D Power, A. Mitra, T. Laumann, A. Z Snyder, B. Schlaggar, S. Petersen, Methods to detect, characterize, and remove motion artifact in resting state fmri 84 (2013).

[29] J. Muschelli, M. B. Nebel, B. S Caffo, A. Barber, J. J Pekar, S. H Mostof-sky, Reduction of motion-related artifacts in resting state fmri using acompcor 96 (2014).

[30] J. A. Frazier, et al., Structural brain magnetic resonance imaging of lim-bic and thalamic volumes in pediatric bipolar disorder., The American journal of psychiatry 162 7 (2005).

[31] J. M. Goldstein, L. J. Seidman, N. Makris, T. Ahern, L. M. O'Brien, V. S. Caviness, D. N. Kennedy, S. V. Faraone, M. T. Tsuang, Hypothala-mic abnormalities in schizophrenia: sex effects and genetic vulnerability, Biol. Psychiatry 61 (2007) 935–945.

[32] N. Makris, J. M. Goldstein, D. Kennedy, S. M. Hodge, V. S. Caviness, S. V. Faraone, M. T. Tsuang, L. J. Seidman, Decreased volume of left and total anterior insular lobule in schizophrenia, Schizophr. Res. 83 (2006) 155–171.

[33] C. D. Smyser, N. U. Dosenbach, T. A. Smyser, A. Z. Snyder, C. E. Rogers, T. E. Inder, B. L. Schlaggar, J. J. Neil, Prediction of brain maturity in infants using machine-learning algorithms, Neuroimage 136 (2016) 1–9.

[34] R. S. Desikan, F. Segonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, R. J. Killiany, An automated labeling system for subdi-viding the human cerebral cortex on MRI scans into gyral based regions of interest, Neuroimage 31 (2006) 968–980.

[35] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, M. Joliot, Automated anatomi-cal labeling of activations in SPM using a macroscopic anatomical par-cellation of the MNI MRI single-subject brain, Neuroimage 15 (2002) 273–289.

[36] C. Cameron, G. James, P. Holtzheimer, X. Hu, M. HS., A whole brain fmri atlas generated via spatially constrained spectral clustering, Human Brain Mapping 33 (2011) 1914–1928.

[37] J. L. Lancaster, M. G. Woldorff, L. M. Parsons, M. Liotti, C. S. Freitas, L. Rainey, P. V. Kochunov, D. Nickerson, S. A. Mikiten, P. T. Fox,

Automated talairach atlas labels for functional brain mapping, Human Brain Mapping 10 (2000) 120–131.

[38] S. B. Eickhoff, K. E. Stephan, H. Mohlberg, C. Grefkes, G. R. Fink, K. Amunts, K. Zilles, A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data, Neuroimage 25 (2005) 1325–1335.

[39] M. D. Schirmer, Developing Brain Connectivity - Effects of Parcellation Scale on Network Analysis in Neonates (Doctoral dissertation, King's College London) (2015).

[40] K. Simonyan, A. Vedaldi, A. Zisserman, Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps, ArXiv e-prints (2013).

[41] A. V. Utevsky, D. V. Smith, S. A. Huettel, Precuneus is a functional core of the default-mode network, Journal of Neuroscience 34 (2014) 932–940.

[42] T. Watanabe, N. Yahata, O. Abe, H. Kuwabara, H. Inoue, Y. Takano, N. Iwashiro, T. Natsubori, Y. Aoki, H. Takao, et al., Diminished medial prefrontal activity behind autistic social judgments of incongruent information, PloS one 7 (2012) e39561.

[43] H. Koshino, P. A. Carpenter, N. J. Minshew, V. L. Cherkassky, T. A. Keller, M. A. Just, Functional connectivity in an fmri working memory task in high-functioning autism, Neuroimage 24 (2005) 810–821.

[44] P. A. Reuter-Lorenz, J. Jonides, E. E. Smith, A. Hartley, A. Miller, C. Marshuetz, R. A. Koeppe, Age differences in the frontal lateralization of verbal and spatial working memory revealed by pet, Journal of cognitive neuroscience 12 (2000) 174–187.

[45] B. Da Mota, V. Fritsch, G. Varoquaux, V. Frouin, J. Poline, T. B., Enhancing the reproducibility of group analysis with randomized brain parcellations, in: Medical Image Computing and Computer-Assisted Intervention - MICCAI 2013. Lecture Notes in Computer Science, vol 8150. Springer, Berlin, Heidelberg.

[46] A. Abraham, M. Milham, A. Di Martino, R. Cameron Craddock, D. Samaras, B. Thirion, G. Varoquaux, Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example (2016).

[47] T. D. Satterthwaite, D. H. Wolf, J. Loughead, K. Ruparel, M. A. Elliott, H. Hakonarson, R. C. Gur, R. E. Gur, Impact of in-scanner head motion on multiple measures of functional connectivity: Relevance for studies of neurodevelopment in youth, NeuroImage 60 (2012) 623 – 632.

[48] D. Fair, J. Nigg, S. Iyer, D. Bathula, K. Mills, N. Dosenbach, B. Schlaggar, M. Mennes, D. Gutman, S. Bangaru, J. Buitelaar, D. Dickstein, A. Di Martino, D. Kennedy, C. Kelly, B. Luna, J. Schweitzer, K. Velanova, Y.-F. Wang, S. Mostofsky, F. Castellanos, M. Milham, Distinct neural signatures detected for adhd subtypes after controlling for micromovements in resting state functional connectivity mri data, Frontiers in Systems Neuroscience 6 (2013) 80.

[49] K. R. Van Dijk, M. R. Sabuncu, R. L. Buckner, The influence of head motion on intrinsic functional connectivity mri, Neuroimage 59 (2012) 431–438.

[50] J. D. Power, A. Mitra, T. O. Laumann, A. Z. Snyder, B. L. Schlaggar, S. E. Petersen, Methods to detect, characterize, and remove motion artifact in resting state fMRI, NeuroImage 84 (2014) 320–341.

[51] P. Hagmann, L. Cammoun, X. Gigandet, R. Meuli, C. J. Honey, V. J. Wedeen, O. Sporns, Mapping the structural core of human cerebral cortex, PLOS Biology 6 (2008) 1–15.

[52] L. M. Zintgraf, T. S. Cohen, T. Adel, M. Welling, Visualizing deep neural network decisions: Prediction difference analysis, CoRR abs/1702.04595 (2017).

[53] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, D. Batra, Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization, CoRR abs/1610.02391 (2016).

7. Supplementary Material

7.1. Atlas Summary

Atlas	# of ROIs	Total Vol.	Median Vol.($\pm$ std)	Min Vol.	Max Vol.
TT	97	1656.34	12.5 ($\pm$16.02)	0.03	69.71
HO	111	1611.39	10.04 ($\pm$15.26)	0.05	97.33
EZ	116	1941.65	14.11 ($\pm$11.96)	0.97	56.35
AAL	116	1843.10	13.78 ($\pm$11.05)	1.35	53.33
DOS160	160	82.05	0.51 ($\pm$0.04)	0.03	0.51
CC200	200	1172.15	5.83 ($\pm$1.26)	1.81	9.96
CC400	400	1172.15	2.97 ($\pm$0.68)	0.76	5.35

Table 4: Summary descriptors of ROIs in individual atlases. All volumes are in cm^3.

7.2. Linear Classifiers

7.2.1. Ridge Classifier

Given feature vectors x_i for n subjects and the corresponding prediction variables denoted by y_i, we approximate the fit using a linear regression model. An L_2 regularization for the weights ($\mathbf{w}$) is added to the mean squared error to yield the following loss function of ridge regression:

$$\mathcal{L}_{\mathcal{R}} = \|Xw - y\|^2 + \alpha \|w\|^2 \tag{1}$$

During classification, the output labels $\mathbf{y}$ are encoded as $\pm$ 1 for the two output categories to minimize the above loss.

7.2.2. Support Vector Machines

(a) Classification

Support Vector Machine Classifiers optimize for a hyperplane with maximum margin between the output classes. This results in a decision function of the form, $f(\mathrm{x}) = \mathbf{sign}(w^T x + b)$. The weights $\{\mathbf{w}, \mathrm{b}\}$ are obtained by minimizing the following convex loss function consisting of a data loss component ($\mathcal{L}_{\mathcal{D}}$) and a regularization loss for the weights ($\mathcal{L}_{\mathcal{W}}$),

$$\mathcal{L}_{\mathcal{SVC}} = C\mathcal{L}_{\mathcal{D}} + \mathcal{L}_{\mathcal{W}} \tag{2}$$

$\mathcal{L}_{\mathcal{D}}$ is modeled using a hinge loss function, $\sum_{i=1}^{n} \max(0, 1 - y_i(w^T x_i + b))$ over all n training samples $\{(x_1, y_1),...,(x_n, y_n)\}$. $\mathcal{L}_{\mathcal{W}}$ is modeled using a Euclidean

norm, i.e., $\|w\|^2$. Here, C is a tuning parameter that controls the trade-off between regularization and data loss.

(b) Regression

The ϵ-Support Vector Regression (SVR) scheme optimizes for a decision function of the form, $f(\mathrm{x})=w^T x + b$, that has at most ϵ deviation from the true prediction variables y (allowing for errors when the problem is infeasible). The loss function ($\mathcal{L}_{\mathcal{SVR}}$) can be formulated as,

$$\mathcal{L}_{\mathcal{SVR}} = C\mathcal{L}_\epsilon + \mathcal{L}_{\mathcal{W}} \tag{3}$$

$\mathcal{L}_\epsilon$ is traditionally referred to as the ϵ-insensitive loss function, and is formulated as $\sum_{i=1}^{n}\max(0, |w^T x_i + b - y_i| - \epsilon)$ over all n training samples $\{(x_1,y_1),...,(x_n,y_n)\}$. The regularization term ($\mathcal{L}_{\mathcal{W}}$) is modeled using a Euclidean norm, i.e., $\|w\|^2$. The tuning parameter C controls the trade-off between the regularization (i.e., the flatness of the decision function) and the amount up to which deviations beyond ϵ are tolerated.

Both the classification and regression problems yield weights **w** that can be represented completely as a linear combination of the training inputs x_i. Thus, **w** is represented as $\sum_{i=1}^{n}\alpha_i x_i$, and the decision function becomes $f(\mathrm{x})=\sum_{i=1}^{n}\alpha_i x_i^T x + b$. This makes it easier to extend SVMs for non-linear decision functions using the kernel technique, i.e., by applying transformations $\phi(\mathrm{x})$ that map x to a high-dimensional space and replacing the inner product $\langle x_i, x\rangle$ with the kernel $K(x_i, x)=\langle\phi(x_i), \phi(x)\rangle$. For our experiments, we observed that the radial basis function kernel, $K(x_i, x) =\exp(\text{-}\frac{\|x_i-x\|^2}{2\sigma^2})$, yields the best results among linear, sigmoid and polynomial kernels up to degree 4.

7.3. ABIDE-I cross-validation results

In order to ensure a fair comparison with other studies that report 10-fold cross-validation performance on ABIDE-I, we report the performance obtained using our benchmark and proposed models (along with the ensemble learning strategy) for both stochastic parcellations and atlases in the form of kernel density plots (Figure S.1). Clearly, the results and conclusions on ABIDE-I remain consistent with ABIDE-II, with the 3D-CNN ensemble strategy outperforming all the baseline methods.

7.4. Saliency maps for individual parcellations

Visualizing the saliency maps for models trained on different brain parcellations can reveal interesting differences in the features captured by these

models. We visualized the saliency maps of the 3D-CNN model for individual stochastic parcellations at multiple scales for the task of ASD/HC Classification. As shown in Figure S.2, models trained using distinct parcellation schemes are relying on the same basic underlying connectivity patterns for prediction, with small differences in their information content, that can be utilized efficiently by the ensemble learning scheme. Further, the saliency maps of atlas-based (see Figure S.3) and stochastic parcellation-based models are remarkably similar, suggesting that the connectivity patterns of the same set of voxels are guiding the classifier predictions, irrespective of the precise scheme of ROI extraction.

ASD/Healthy Classification: 10-fold cross-validation on ABIDE-1
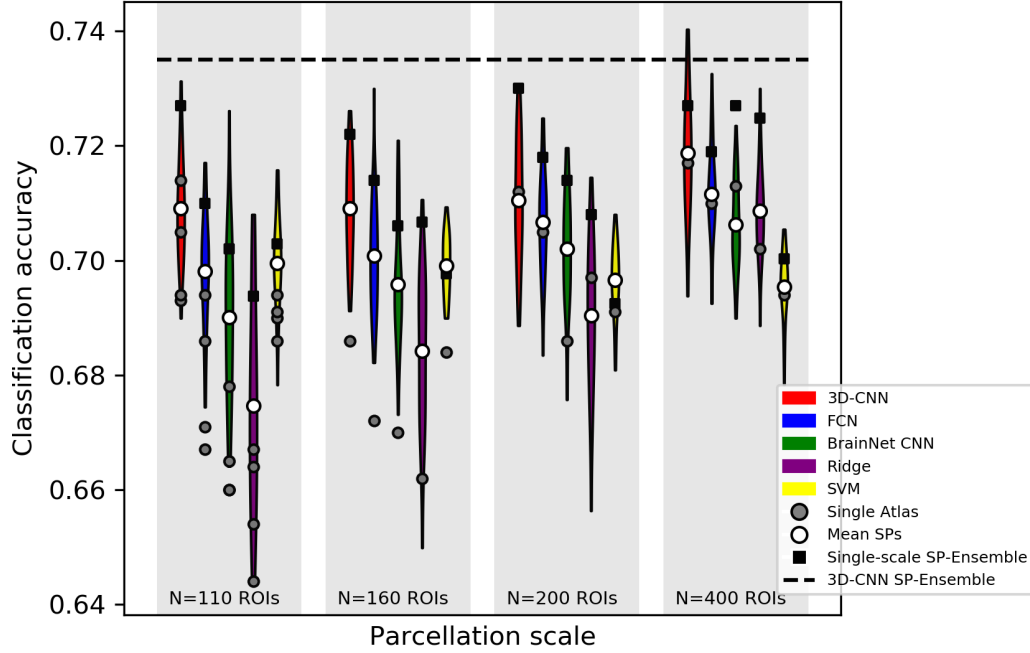
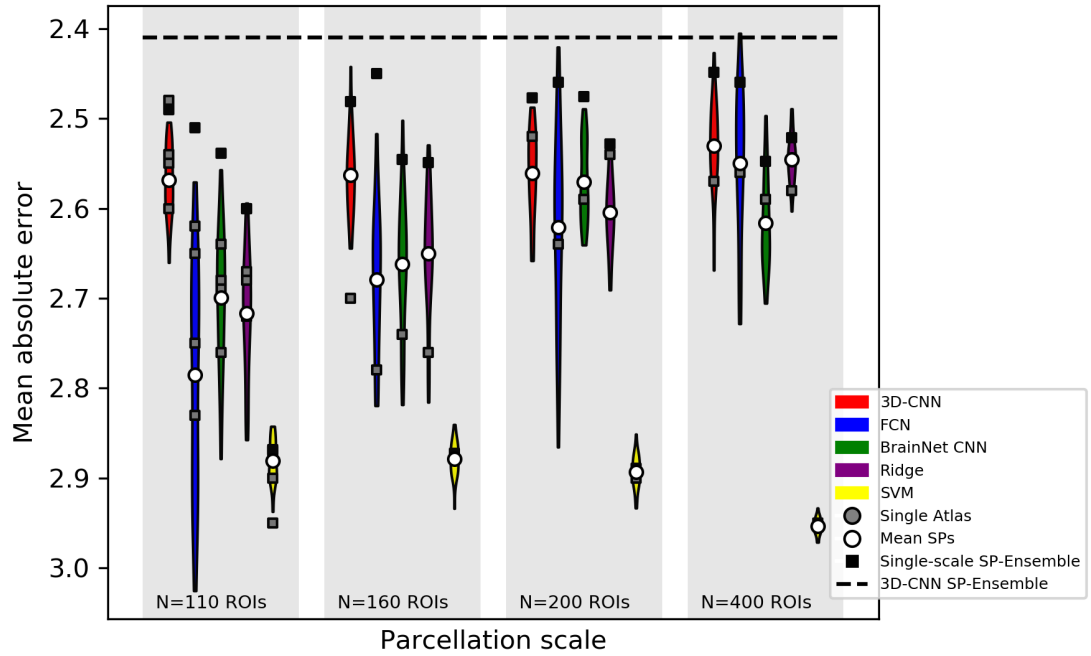Age Prediction: 10-fold cross-validation on ABIDE-1

Figure S.1: Violin plots showing the spread of prediction accuracies/errors for stochastic parcellations at multiple network scales for different classification models. Mean accuracy/error of individual violins is denoted by 'Mean SPs'. Performance of individual atlases is compared with SPs with the closest # of ROIs and is denoted as 'Single Atlas'. Results are computed by 10-fold cross-validation on the entire ABIDE-1 cohort.
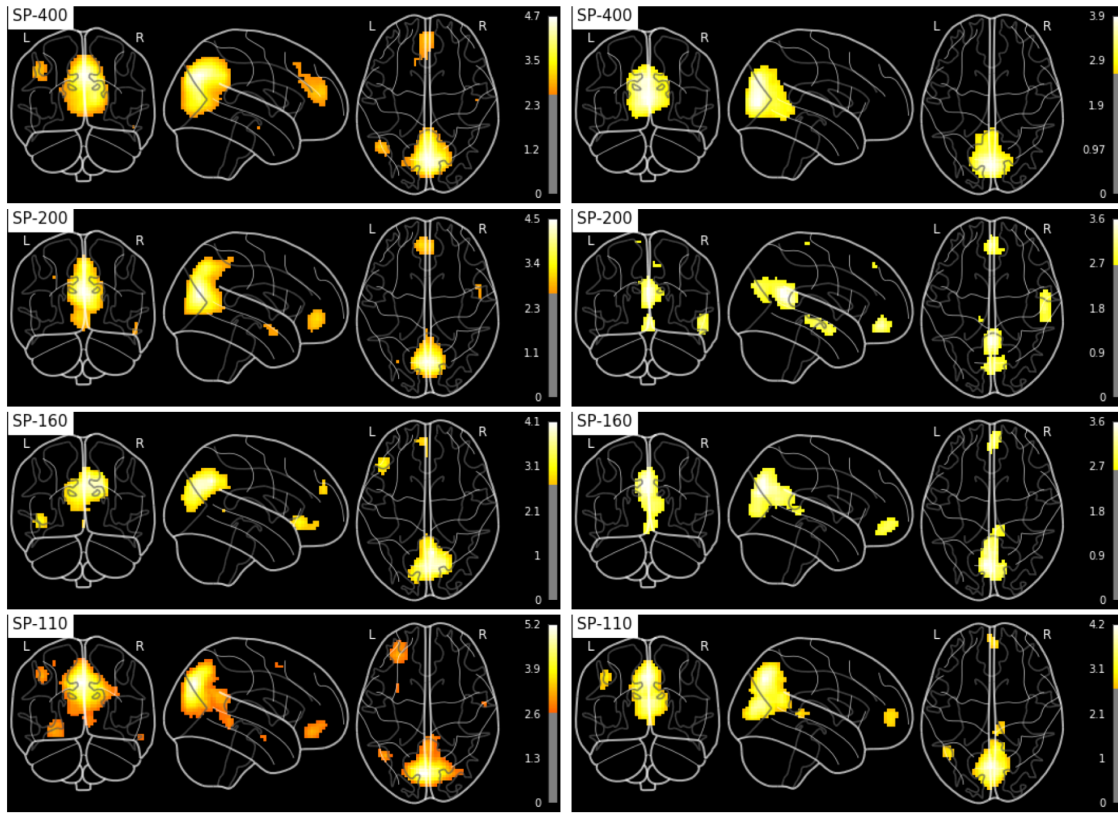
Figure S.2: Saliency maps of trained CNN models for 2 randomly chosen stochastic parcellations at each scale for ASD-HC classification.
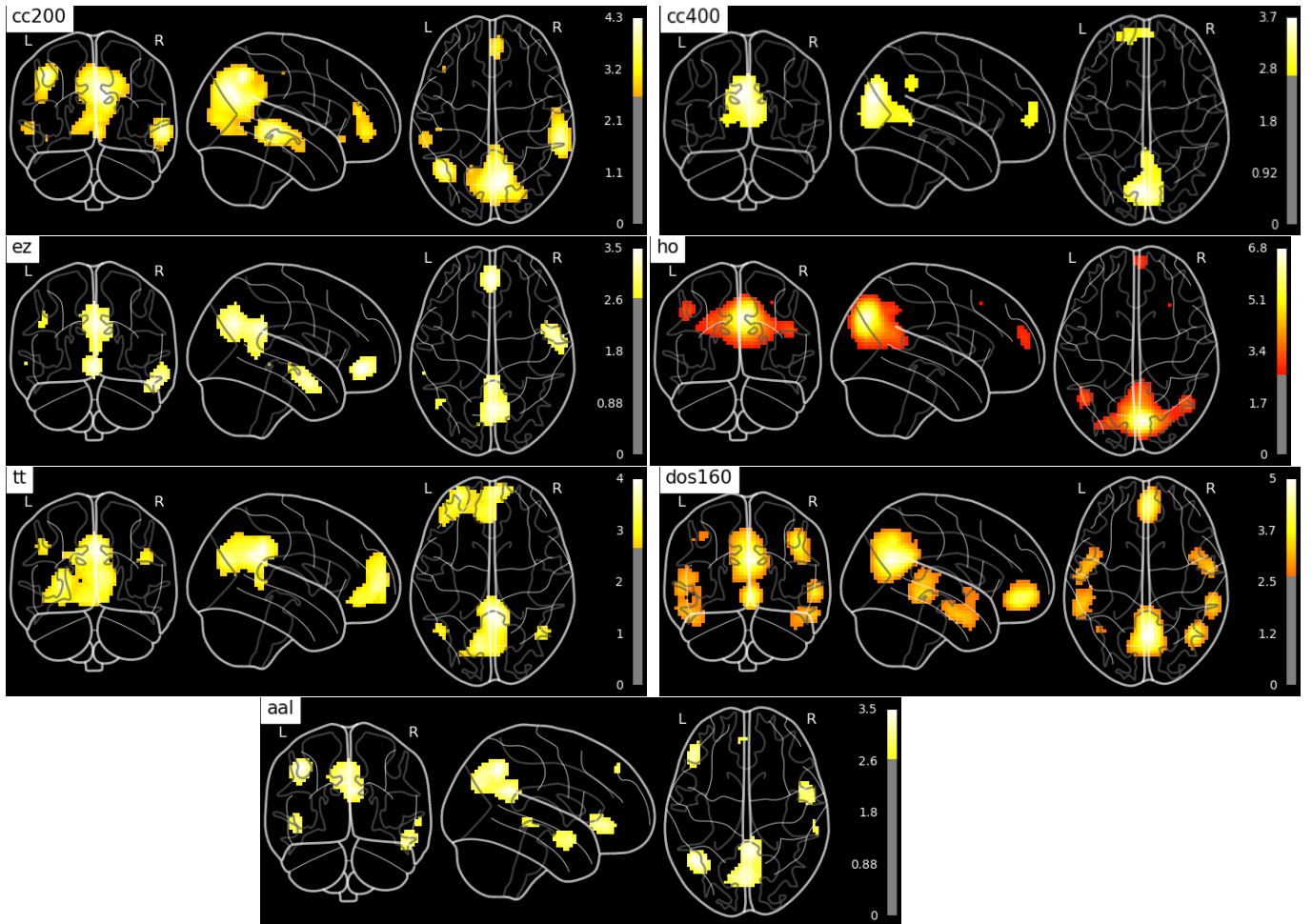
Figure S.3: Saliency maps for atlas-based ASD-HC classifiaction models.