ENSEMBLE DEEP LEARNING ON LARGE, MIXED-SITE FMRI DATASETS IN AUTISM AND OTHER TASKS

MATTHEW LEMING

Department of Psychiatry, University of Cambridge, Herchel Smith Building, Robinson Way,
Cambridge, CB20SZ, UK
E-mail: ml784@cam.ac.uk
www.cam.ac.uk

JUAN MANUEL GÓRRIZ

Department of Signal Theory, Networking and Communications, University of Granada, Avenida del Hospicio,
Granada, 18071, Spain
www.ugr.es

JOHN SUCKLING

Department of Psychiatry, University of Cambridge, Herchel Smith Building, Robinson Way,
Cambridge, CB20SZ, UK
www.cam.ac.uk

Deep learning models for MRI classification face two recurring problems: they are typically limited by low sample size, and are abstracted by their own complexity (the "black box problem"). In this paper, we train a convolutional neural network (CNN) with the largest multi-source, functional MRI (fMRI) connectomic dataset ever compiled, consisting of 43,858 datapoints. We apply this model to a cross-sectional comparison of autism (ASD) vs typically developing (TD) controls that has proved difficult to characterise with inferential statistics. To contextualise these findings, we additionally perform classifications of gender and task vs rest. Employing class-balancing to build a training set, we trained 3×300 modified CNNs in an ensemble model to classify fMRI connectivity matrices with overall AUROCs of 0.6774, 0.7680, and 0.9222 for ASD vs TD, gender, and task vs rest, respectively. Additionally, we aim to address the black box problem in this context using two visualization methods. First, class activation maps show which functional connections of the brain our models focus on when performing classification. Second, by analyzing maximal activations of the hidden layers, we were also able to explore how the model organizes a large and mixed-centre dataset, finding that it dedicates specific areas of its hidden layers to processing different covariates of data (depending on the independent variable analyzed), and other areas to mix data from different sources. Our study finds that deep learning models that distinguish ASD from TD controls focus broadly on temporal and cerebellar connections, with a particularly high focus on the right caudate nucleus and paracentral sulcus.

Keywords: Autism; Big data; Functional connectivity; Deep learning.

1. Introduction

The characterization of brain differences in autism spectrum disorder (ASD) is an ongoing challenge. Although the consensus is that there are widespread structural and functional differences, the direction and spatial patterns of differences are not reliably observed and overlap with inter-individual variability in the neurotypical population.

Estimates of grey matter volume with voxel-based morphometry (VBM) have been the most com-

monly used methodology to assess brain structure, but have resulted in discrepancies amongst meta-analytic findings,[1–3] at least a partial explanation for which are the small-sample sizes that are a prevalent feature of the primary literature.[4, 5]

To address variations in data acquisition and processing that make between-study comparisons less powerful, publicly available large-sample datasets are now pivotal to imaging research. The ABIDE multi-centre initiative has made available over 2000 images in two releases, but cross-sectional VBM analyses have failed to observe significant differences.[6, 7] Other morphological properties of the cortex may yield greater sensitivity,[8] and recent findings using estimates of cortical thickness from the ENIGMA working group suggest a complex pattern of differences relative to neurotypical controls that varies across the lifespan.[9] Other databases, such as the National Database for ASD Research (NDAR) act as aggregates of MRI data for different smaller-scale studies, though centre differences complicate conventional analyses on these data as a whole.

ASD has been consistently associated with differences in brain function.[10, 11] This is often studied in the context of EEG,[12–15] for which several studies have been conducted to achieve automated diagnosis,[16–21] and fMRI. The measurement of correlation, or "functional connectivity", between time-series of blood oxygenation level dependent (BOLD) endogenous contrast estimated from brain regions during resting wakefulness has been demonstrated as a reproducible measurement on an individual basis.[22] Functional connectivity (FC) matrices are estimates of the connectivities between all brain regions that can be represented as undirected graphs (connectomes) of nodes (brain regions) and edges (connectivity strengths). They show promise in localising characteristic differences for ASD in resting activity to specific large-scale brain networks.[23] Whilst there is cautionary evidence using the ABIDE dataset and others,[24] it would appear that statistically significant differences in connectivity are generally observable, but like measurements of brain structure, are variable in their presentation. With consistent and localised changes remaining elusive, a number of studies have characterised ASD as exhibiting under-connectivity in certain areas of the brain,[25–30] while others show evidence of over-connectivity.[31–36] A recent review[37] posited that ASD is likely a mix of these traits.

In other fields, computing power and access to large datasets have led to a resurgence in the popularity of NNs as a tool for data classification.[38–40] NNs are especially adept at classifying complex data which parametric inferential statistics may fail to fully characterize due to their inherent assumptions. Given that brain function in ASD has been consistently found to be different but in different ways, such a model may be a sensible approach for a comprehensive representation. In parallel, because of their wide applicability in representing complex data such as proteins and social networks, functional connectomes have undergone significant development in terms of global and local topological descriptions. Some recent work has used NNs for processing connectomes, including whole-graph classification, clustering into sub-graphs, and node-wise classification.[41–46] Previous efforts to classify functional connectivity in ASD on smaller datasets have achieved accuracy rates that have been described as "modest to conservatively good",[37] though these methods have had trouble replicating on different data.[47–49] More recently, the application of convolutional CNNs to ABIDE data has achieved achieved 68% to 77.3% classification accuracies.[50–53]

In this article, we leverage publicly available datasets to amass and automatically pre-process a total of 43,838 functional MRIs from nine different collections. To test the application of CNNs to imaging data, we first classify autistic individuals from typically developing (TD) controls. To validate the proposed models, we then classify functional connectivity matrices based on gender and task vs resting state. All classifications were undertaken using a CNN that uniquely encodes multi-layered connectivity matrices, using an original deep learning architecture, partially inspired by Kawahara et al 2017. We opted to use these connectivity matrices as opposed to full fMRI datasets both for memory management purposes (the average fMRI dataset in our collection is 176 MB per file, while the connectivity matrix is just under 500 KB), and for interpretability, as connectivity matrices allow for the direct analysis of both localized areas and connections between areas. Due to the stochastic properties of NNs and set divisions, we use a standard stratified cross-validation strategy, performing each of our tests across 300 independent models using different subsamples and di-

visions of the total dataset. To incentivise the model to classify based on phenotypic differences rather than centre differences, class-balancing techniques across participant age and collection were used when building the training and test sets, and compared against the fully-inclusive samples.

Key outputs of the CNN are class activation maps[54–56] that highlight areas of the connectome the model preferentially focuses on when performing its classification, and activation maximization[57] of a hidden layer that visualizes how the model partitioned the dataset as a whole following classification. We suggest an index to quantify the output of activation maximisation.

In attempting to classify components of this accumulated dataset, we sought to address the following questions: (1) How effective is our machine learning paradigm at classifying FC in ASD, gender, and resting-state/task? (2) Which areas or networks of the brain do models focus on when undertaking classifications? (3) How does the model partition large datasets during different classification tasks? (4) Can the model effectively classify FCs taken from multiple sources without relying explicitly on centre differences to do so?

2. Methods

2.1. *Datasets and preprocessing*

Datasets were acquired from OpenFMRI;[58,59] the Alzheimers Disease Neuroimaging Initiative (ADNI); ABIDE;[60] ABIDE II;[61] the Adolescent Brain Cognitive Development (ABCD) Study;[62] the NIMH Data Archive, including the Research Domain Criteria Database (RDoCdb), the National Database for Clinical Trials (NDCT), and, predominantly, the National Database for Autism Research (NDAR);[63] the 1000 Functional Connectomes Project;[64] the International Consortium for Brain Mapping database (ICBM); and the UK Biobank; we refer to each of these sets as *collections*. OpenFMRI, NDAR, ICBM, and the 1000 Functional Connectomes Project are collections that comprise different datasets submitted from unrelated research groups. ADNI, ABIDE, ABIDE II, ABCD, and the UK Biobank are collections that were acquired as part of a larger research initiative.

These data were pre-processed using the fMRI Signal Processing Toolbox (SPT). Following skull-stripping, motion correction was accomplished using SpeedyPP version 2.0, which utilized AFNI tools and wavelet despiking,[65,66] with a low-bandpass filter of 0.01Hz, in addition to motion and motion derivative regression. Both functional and structural datasets were non-linearly registered to MNI space and parcellated using the 116-area automated anatomical labeling (AAL) template,[67] which includes subcortical regions. Extracted time series were the means of each AAL region. Each dataset was transformed into N $4 \times 116 \times 116$ connectivity matrices, using edges weighted by the Pearson correlation of the wavelet coefficients of the pre-processed time-series in each of four frequency scales: 0.1-0.2 Hz, 0.05-0.1 Hz, 0.03-0.05 Hz, and 0.01-0.03 Hz. Wavelet correlation estimates were adjusted from TR rates to equalize the frequency ranges across different collections. Pre-processing was accomplished on a computing cluster over a period of several weeks. Due to the volume of datasets, individualized quality control was not possible. The porportion of datasets failing pre-processing varied by collection.

Across all collections, 70,284 potential datasets were identified of which 67,396 contained suitable functional and structural datasets. Of these, 52,396 succeeded pre-processing to parcellation. However, datasets with regional dropout of greater than 10% were omitted from the analyses, and redundant datasets across collections were also discarded along with those data with a TR outside of the desired range. In total, 43,838 connectomes from 17,614 unique participants were available for analysis with the NN. Multiple instances of connectomes from the same individuals were used, though they were not shared between the training, validation, and test sets. The numbers of participants, total numbers of datasets used as well as phenotypic distributions, are shown in Table 1.

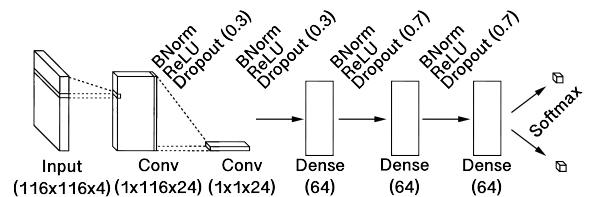### 2.2. *Neural Network Model and Training*

Figure 1. The structure of the neural network. These were applied in an ensemble model, so the outputs of 300

Table 1.   Average populations present for successfully-preprocessed datasets. Some datasets were not labeled with respect to one or more covariates, so counts may not sum to the listed total.

					Age				Sex		Disorders
Collection	Subjs	Conns	Rest	Task	Min	Max	Mean	Stddev	F	M	Autism
1000 FC	764	764	764	0	7.88	85.00	25.76	10.18	443	321	0
ABCD	1319	9205	4043	5162	0.42	11.08	10.08	0.65	4339	4866	113
Abide	193	193	193	0	9.00	50.00	17.81	6.69	21	172	94
Abide II	720	761	761	0	5.22	55.00	14.44	7.45	174	587	375
ADNI	141	261	261	0	56.00	95.00	73.57	7.32	146	115	0
BioBank	11811	16970	9937	7033	40.00	70.00	55.23	7.51	8752	8218	8
ICBM	112	381	29	352	19.00	74.00	43.53	14.83	188	193	0
NDAR	1123	8569	5952	2617	0.25	55.83	18.65	7.82	4165	4404	994
Open fMRI	1443	6655	1169	5486	5.89	78.00	27.22	10.40	2768	3133	127
All	17614	43838	23109	20650	0.25	95.00	33.05	20.68	20996	22009	1711

independently-trained neural networks were averaged in a cross-validation scheme.

The data used for training and testing the CNN were $4 \times 116 \times 116$ (4 wavelet scales and 116 nodes) symmetric FC (wavelet coefficient correlation) matrices, with values linearly scaled from [-1.1] to [0,1] for easier use in a NN.

To classify the data, we employed a CNN with vertical convolutional filters on the first layer followed by horizontal convolutional filters on the second layer, effectively reducing the matrices to single values to allow the network to train on connectivity matrices (Figure 1). This approach was partially inspired by the cross-shaped filters described in Kawahara et al 2017, though previous tests with that architecture resulted in a number of failed models with no apparent increase in accuracy over the simpler architecture proposed here. We implemented this architecture using Keras,[68] a popular machine learning library, leveraging the advantages of supporting software libraries. Additionally, this implementation includes multiple channels in the inputs, as opposed to single-input connectivity matrices.

The CNN was constructed with: 24 edge-to-node vertical convolutional filters; 24 node-to-graph horizontal convolutional filters; 3 fully-connected layers, each with 64 nodes; and a final softmax layer. Separating each layer were batch normalization, rectified linear unit (ReLU), and dropout layers, with the dropout being 0.3 in the convolutional layers and 0.7 in the dense layers. The layer structures and ordering followed the advice offered in 69. Specifications are shown in Figure 1. No pooling layers were

used, and all strides were of length 1. The model was trained using an Adam optimizer with batch sizes of 64. Otherwise, Keras defaults were used. Models were trained for 200 epochs, and the epoch with the highest validation accuracy was selected.

To obtain a reliable average, we trained 300 models independently for each classification, which were then combined in an ensemble model. In each training instance, a subset of the total available data was taken. A holdout test and validation set were not used,[70] but instead a division of the data was performed for each model in a stratified cross-validation schema, subject to the rules detailed below.

2.3. *Set division*

Data were divided into three sets: a training set, comprising of two-thirds of the data and used to train the model; a validation set, comprising of one-sixth of the data and used to select the epoch at which training stopped; and a test set, used to assess the trained classifier performance, comprising of one-sixth of the data. The approximate total number of images used by each model was 10,000 for the gender and resting-state classification, and 4000 (limited by sample size) for the ASD classification. For all classifications, balancing was used such that each class comprised approximately half of the datasets. To account for covariates, classes were additionally balanced such that the distributions of different collections and ages were equal between classes. For collection balancing, equal numbers of datasets were used from each collections. For continuous age values, distributions of age between classes were made to fail a Mann-Whitney U-test, with p¿0.05. We used stan-

dard stratified cross-validation rather than a holdout division across the 300 runs.

Because of the collection balancing procedure, many data were excluded from certain classification tasks; for instance, as BioBank only included eight subjects with ASD. Due to the class balancing, set divisions were not precise in each instance.

2.4. *Test set evaluation*

2.4.1. *Inter-data classification*

Following the training of the models, the accuracy and the area under the receiver operating characteristic curve (AUROC) were calculated as measures of machine learning performance on the test set. This was to determine if one group in the classification outperformed the other in training leading to a biasing of the overall accuracy.

2.4.2. *Activation Maximization*

Activation maximization[57] is a technique to determine the maximally activated hidden units in response to the test set of the CNN layers following training. Activation maximization was applied to the 116×24 second layer of our network (Figure 1) as this convolutional layer acts as a bottleneck, and is thus easier to interpret and visualize. This layer is naturally stratified by 24 *filters*, each with 116 nodes (brain regions). To offset the influence of spurious maximizations, we opted to record the 10 datasets that maximally activated each hidden unit, obtaining their mode with respect to collection, gender, and whether it was task/rest; for example, if six connectomes that maximally activated a unit were from Collection A and four were from Collection B, Collection A would be recorded as maximally activating that hidden unit.

For each covariate, this method yields a 116×24 array of values for each of the 3×300 models. We opted to measure the stratification of the different convolutional filters in our models by measuring whether it was maximally activated primarily by one source of data, or whether it was activated by a mixed population. With this in mind, we calculated for each layer a diversity coefficient, which is 0 if the layer is only maximally activated by one class of data and 1 if it is maximized proportional to the population maximized. Given K possible classes, $F_k, k \in K$ indicating the percentage of each class in a given fil-

ter, and $T_k, k \in K$ indicating the percentage of each class across all filters, we calculated the diversity coefficient for each filter as:

$$D_i = \frac{\tan^{-1}\left(\ln \frac{1 - \sqrt{\sum_{k=1}^{K} F_k^2}}{\sqrt{\sum_{k=1}^{K} \frac{(F_k - T_k)^2}{2}}}\right) + \frac{\pi}{2}}{\pi} \quad (1)$$

Briefly, the justification for this equation is that the summation $\sum_{k=1}^{K} \frac{(F_k - T_k)^2}{2}$ equals 0 if the distribution of the filter's population is equal to the population of the whole layer; that is, the distribution is ideally diverse, and this pulls the logarithm towards $-\infty$, which in turn pulls the inverse tangent function to $\frac{\pi}{2}$. Conversely, $1 - \sqrt{\sum_{k=1}^{K} F_k^2}$ tends towards 0 if the individual layer is only composed of a single class, pulling the inverse tangent towards $-\frac{\pi}{2}$. The diversity coefficient is normalized to be between 0 and 1. Its value is indeterminate if only a single class is present globally.

This equation is a more complex version of other diversity coefficients, such as the Herfindahl-Hirschman or Simpson diversity indices. However, the proposed index better accounts for overall populations in the hidden layer activations and thus makes it easier to compare across different classification tasks and independent variables. While the Herfindahl-Hirschman or Simpson indices both approach their maxima when the measured population is completely homogenous, their lower extrema varies depending on the number of distinct populations present. This is problematic in comparing across indices, because the number of populations varies depending on the application, and assumes that the expected (i.e., most diverse) distribution occurs when different populations are perfectly proportional. The proposed index defines the most diverse population as that which has distributions proportional to the overall population, at which point the index is zero.

In practice, low diversity coefficients indicate that the ensemble models stratified data by the covariate. This allows us to measure the degree to which individual covariates (such as collection) were taken into account by the CNNs. We found the diversity coefficient of each of the 24 filters of our hidden, 116×24, convolutional layers, then sorted these values to show which filters were primarily activated by a few covariates and which were activated maximally by many covariates.

Table 2.   The ensemble and averaged AUROCS and accuracies   for   300   models.

	Autism	Gender	Rest v Task
Ensemble AUROC	0.6774	0.7680	0.9222
Ensemble Acc.	67.0253%	69.7063%	85.1996%
Average AUROC	0.6133	0.6858	0.9231
Average Acc.	57.1150%	63.3398%	84.3153%

2.4.3.   *Class Activation Maps*

We used class activation maps (CAMs)[54–56] and a previous Keras implementation[71] to display parts of the connectivity matrix the CNN emphasised in its classification of the test sets. CAMs operate by taking the derivative of the CNN classification function (approximated as a first-order Taylor expansion, estimated via back-propagation) with respect to an input matrix, with the output being the same dimensions as the input.[54, 56] While class activation maps were originally proposed in,[54] they were improved to the commonly-used method presented in,[56] known as 'Gradient Class Activation Maps (Grad-CAMs). CAMs are particularly advantageous when applied to connectivity matrices, because unlike typical 2D images, these matrices are spatially static (i.e. each part of the matrix represents the same connection in the brain, across all datasets). Thus, global tendencies of the model can be visualized by averaging many CAMs. CAMs for each connectivity matrix were averaged, maximised across the four wavelet frequency domains, and displayed to show which aspects of the connectome the CNN focused. To simplify the analysis, CAMs were taken with respect to the input data's predicted output, rather than two output classes.

2.5.   *Experiments*

We performed the classification on class- and age-balanced datasets that then classified based on gender, task vs rest, and ASD vs TD controls in separate analyses. We then analysed the averaged CAMs with respect to their output prediction. We also recorded the diversity coefficient with respect to gender, collection, and rest v task.
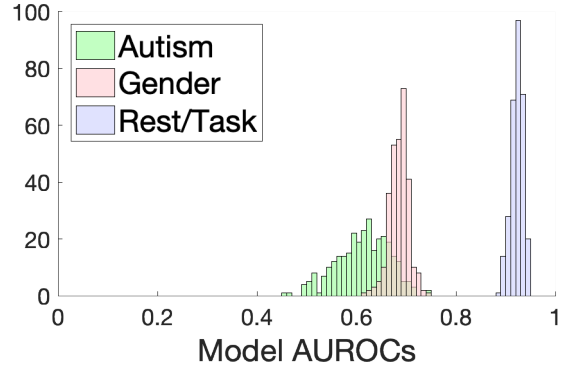

Figure 2.   Histograms of all AUROCS for 300 independent models, using different, stratified samples of the whole dataset.

3.   Results

Table 2 shows the accuracies for the 300 models tested. The AUROCs for the individual models, across all data (Figure 2) were averaged to give 0.6858, 0.9231, and 0.6133 for gender, task vs rest, and ASD vs TD classifications, respectively, while the average accuracies were 63.33%, 84.31%, and 57.11%. In nearly all cases, however, as shown in Table 2, the ensemble AUROC and accuracies were substantially higher. The ROC of ensemble models with respect to collections are shown in Figures 3C, 4C, and 5C.

The results in Figures 3B, 4B, and 5B display the histogram of diversity indices across all models' activation maximisation values. This indicates the tendency of models to use particular filters to sequester data by different covariates, especially if it were attempting to classify by that variable; thus, a diversity index of 0 indicates that all nodes within a particular filter were maximally activated from one or a small number of collections (i.e., BioBank or Open fMRI). The covariates measured are gender, rest/task, and collection site; ASD was not included as a covariate because of the relatively small percentage of ASD data overall.

The diversity index of the activation maximization of the second hidden layer revealed that filters we in many cases sorted into two distinct groups, as shown by peaks on the lower and upper end of histograms in Figures 3B, 4B, and 5B: stratified layers (i.e., with a diversity index close to 0), which were wholly maximally activated by one type of dataset, and mixed layers (i.e., with a diversity index close to 1), which integrated data from different sources.
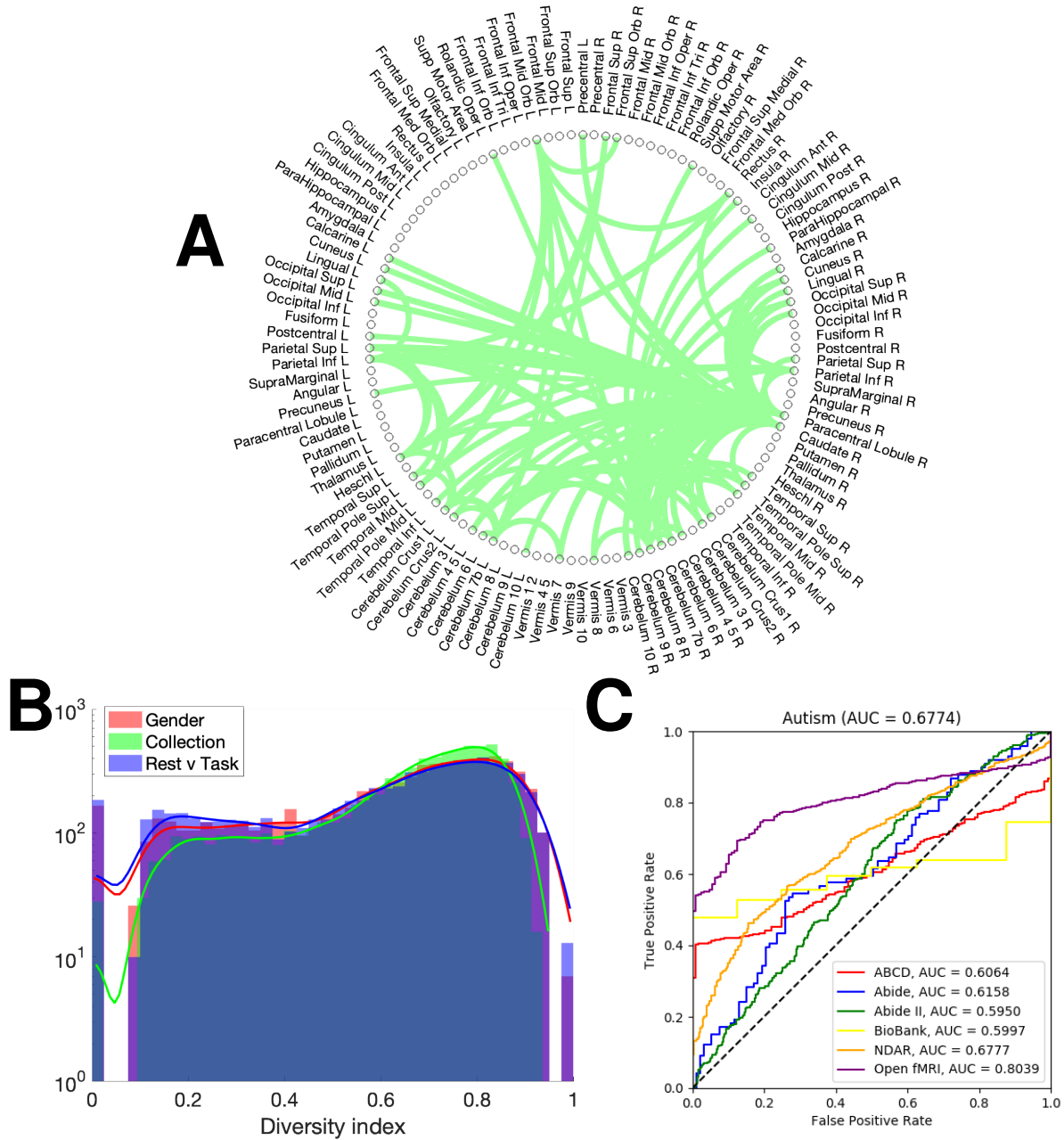
Figure 3.   Results for autism classification. (A) The 100 strongest connections of the mean class activation maps, with the maximum value taken across wavelet correlations. (B) The distribution of the diversity index of maximal activations across all filters over 300 models, showing how much filters in general were dedicated to particular phenotypes. (C) The overall classification AUROC and the AUROC of individual data collections in the model, showing the overall and relative success of the model.

While gender and task vs rest each had a proportion of their filters wholly activated by a single collection, the majority of filters were activated by a variety of different collections, indicating the effective synthesis of data from different sources. ASD, however, had a large proportion of data with a diversity in-dex close to zero; this is expected for the gender and resting-state covariates, given that the datasets were mainly from males, but the low diversity indices for collection indicates that ASD classification models sequestered data based on collection, and thus many datasets were considered independently.

3.1.   *ASD vs TD Controls*

With class balancing, the ensemble performance for ASD v TD controls across test sets was AUROC=0.6774 (Figure 3). ASD classifications were highly dependent on the collection used, although the final AUROCs were above chance for all collections. Class balancing was particularly necessary for this scheme, as data from autistic individuals comprised less than 10% overall.

As stated above, activation maximization saw high stratification with regards to gender and resting-state (Figure 3B). Collection also saw a mix of filters that were both highly stratified and highly diverse, indicating the dual use of convolutional filters. Given the phenotypic differences in our ASD datasets (with ABCD consisting largely of children and ABIDE adolescents, for instance), it is likely that the models considered parts of them independently during classification.

Class activation was strongest for ASD in the limbic system, cerebellum, temporal lobe, and frontal middle orbital lobe, but overwhelmingly emphasised in the right caudate nucleus and paracentral lobule (Figure 3A). Findings of the caudate nucleus are consistent with historical findings in developmental ASD,[72] with both aberrant FC frequently associated with that area and the presence of volume differences.[73–80]

3.2.   *Gender*

The ensemble classification of gender yielded 0.7680 AUROC, with comparable AUROCs across different collections (Figure 4C).

On average, CAMs in gender classifications showed more differences around areas in the corpus callosum and the frontal lobe (especially the medial left frontal lobe), as well as parietal areas, with very few subcortical differences (Figure 4A).

In activation maximization (Figure 4B), most of the filters mixed data from different genders and rest/task. A proportion were maximally activated by individual collections, but for the most part, this was mixed as well. Among the three classification tasks in this study, gender integrated the most data from different sources. As gender distributions are likely the most homogenous variable tracked across datasets (with the exception of ABIDE I and II), the stratification with respect to individual collections was

appropriately lower than expected when classifying other variables.

3.3.   *Task vs Rest*

Task v rest classification had an ensemble classification of AUROC=0.9222 (Figure 5C), by far the highest of any classification task. BioBank rest/task classification had nearly perfect classification, while other collections that contributed substantial amounts to both resting-state and task participants, that is, NDAR, ABCD, and Open fMRI, had comparable performance. The CAM focused on the default mode network, largely in the left hemisphere, and its connection to the right frontal medial orbital area. The highly emphasised areas include the supplementary motor area, the left parietal lobe, the bilateral middle and inferior occipital lobe, the left precentral gyrus, and the bilateral thalamus representing the wide range of areas activated in task fMRI.

In activation maximization, stratification was found with respect to task (the target covariate), somewhat on collection, and very little with respect to gender. A degree of collection stratification may be expected due to the different tasks found in different collections; for instance, BioBank consisted almost entirely of an emotional faces recognition task, while Open fMRI contains a medley of different tasks.

4.   Discussion

This work describes how large and diverse imaging data might be analyzed by deep learning models, encouraging the aggregation of publicly available collections. Data were partitioned based on clear and logical features of the images and, even with imperfect classification accuracies, deep learning models were capable of recognizing complex patterns in large datasets, many consistent with previous work.

The neuroscientific objective of this study was to use available imaging data with deep learning to describe the pattern of functional brain changes that distinguishes ASD from TD. With the absence of any gold standard in this cross-sectional comparison,we also undertook classifications of gender and rest v task, which have more secure, robust findings in the extant literature to confirm the veracity of the developed methods.

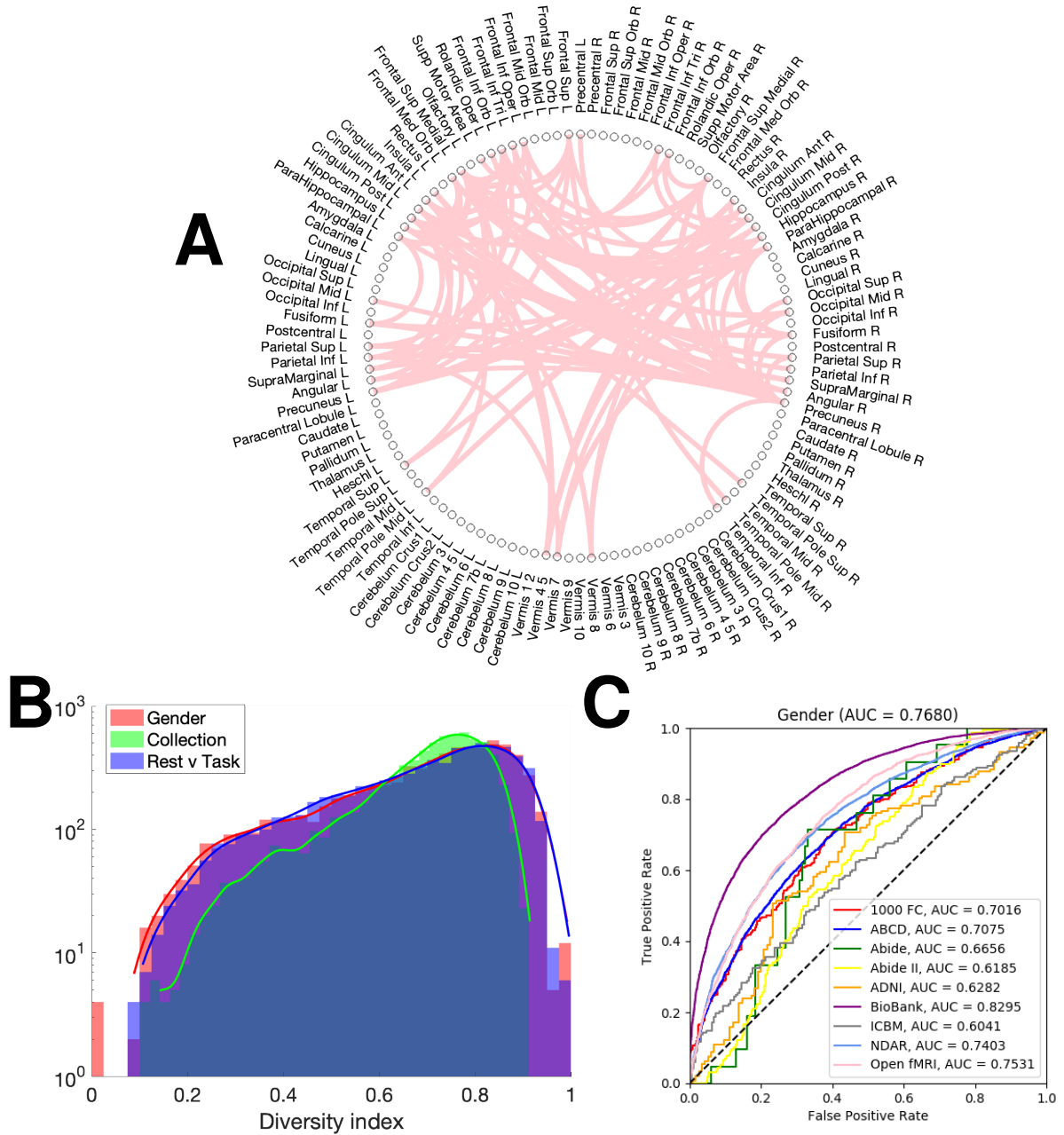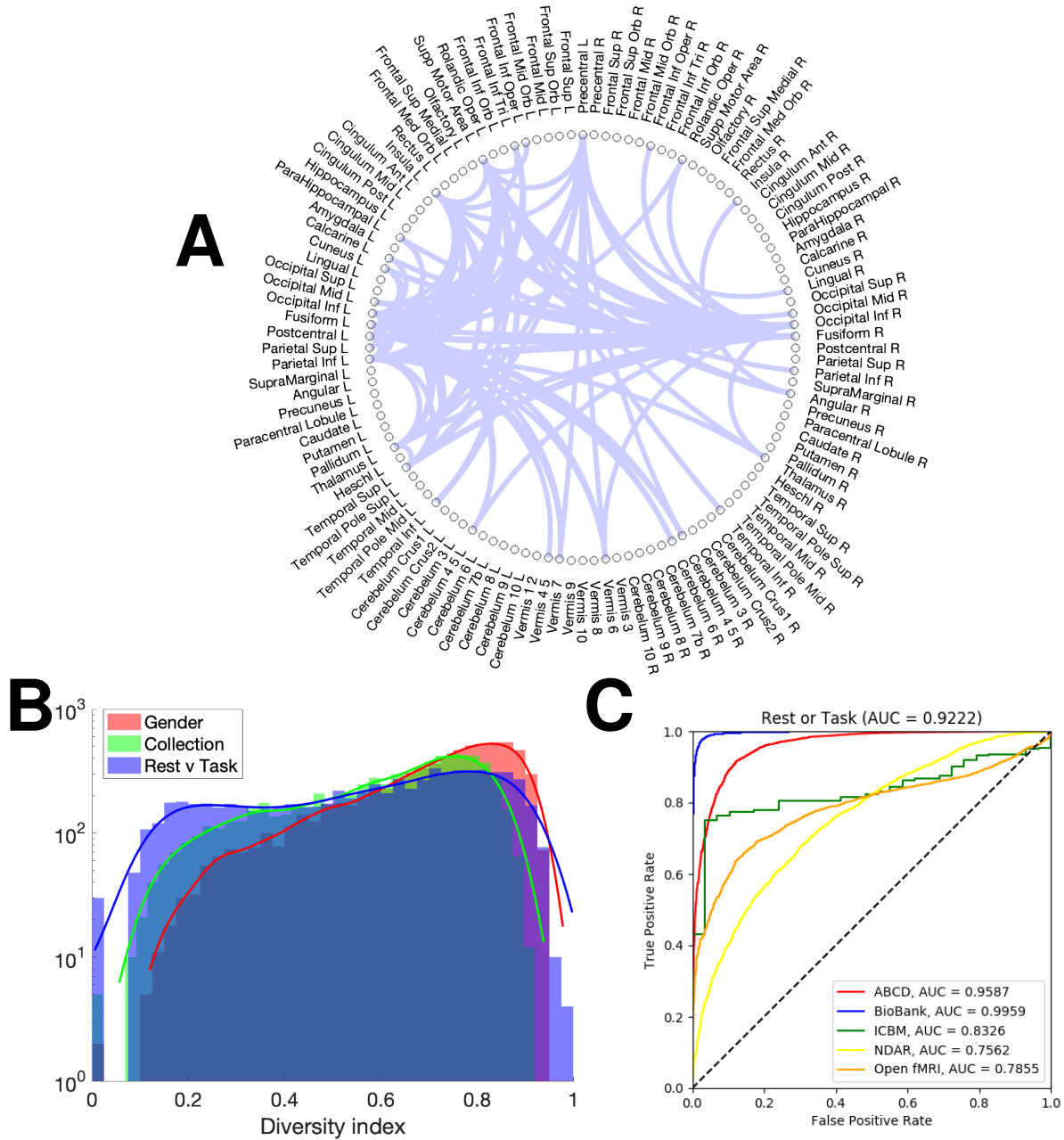We used CAMs[54, 56] to identify connections and

Figure 4. Results for gender classification. (A) The 100 strongest connections of the mean class activation maps, with the maximum value taken across wavelet correlations. (B) The distribution of the diversity index of maximal activations across all filters over 300 models, showing how much filters in general were dedicated to particular phenotypes. (C) The overall classification AUROC and the AUROC of individual data collections in the model, showing the overall and relative success of the model.

areas that had a pronounced influence on the classifications by the model. This method has previously been used in deep learning on functional connectivity[53,55] as an effective way of dissecting NNs. However, a caveat to this is that CAMs, while indicators of areas of importance in the data, may not give a complete depiction of its distinguishing features. Without further tests, CAMs cannot indicate whether a particular set of edges is over-connected or under-connected, or whether the areas of high class activation are independent or components of a more complex pattern.

Figure 5.   Results for resting-state/task classification. (A) The 100 strongest connections of the mean class activation maps, with the maximum value taken across wavelet correlations. (B) The distribution of the diversity index of maximal activations across all filters over 300 models, showing how much filters in general were dedicated to particular phenotypes. (C) The overall classification AUROC and the AUROC of individual data collections in the model, showing the overall and relative success of the model.

When classifying gender, the model was influenced by diffuse areas connected to the frontal lobe (Figure 4A). This is consistent with previous findings in gender comparisons of functional imaging, which did not find differences in brain activity in specific areas, but rather differences in local FC over large areas of the cortex.[81]

Task vs rest FC classifications, as expected, identified the major components of the well-known default mode network[82] (Figure 5A), a set of bilateral and symmetric regions that is suppressed during exogenous stimulation,[83] as well as visual process-

ing areas (the occipital lobe) and the supplementary motor area. Together with the comparison of gender, the confirmation of the results with those expected from the extant literature give confidence for accurate classification by the CNN as well as the specificity of the visualization method used.

The paracentral lobule and right caudate nucleus, as well as connections to the cerebellum and vermis, were identified as salient to the comparison of the ASD vs TD (Figure 3). This finding is largely substantiated by previous studies that have found both FC and volume differences between autistic and healthy individuals in the caudate nucleus,[73–79] though these studies disagree on the exact nature of those differences.[80] Much of the literature on functional connectivity in ASD, however, concerns network-wide differences[37] rather than localized differences captured by the CAMs.

Another of the key methods we used to interrogate the results from our deep learning model was activation maximization. Previously, activation maximization has been used for intuiting the internal configuration of NNs rather than for quantitative interpretation,[57] which has never been tried, especially across many different independent models. Many of the filters in our models were wholly activated by datasets from a single group, while others utilised a mixture of datasets. We sought to quantify this effect through a diversity index leading to two general observations: first, across models, a few filters were entirely activated by a single collection (i.e., had a diversity index of 0), though which collection remained inconsistent, and was not apparently proportional to the amount of data contributed by that particular dataset; Second, across models, the diversity index was not normally distributed but often had two peaks, one at the low end of the spectrum (indicating stratification of the filters) and one at the high end (indicating a highly diverse, or close to random, distribution of the filter). In ASD, a disproportionately high number of filters were activated by a single collection, indicating that the NN split data internally more than other classification tasks.

In ASD, model accuracy was lower compared to the highest rates reported in literature,[51–53] although this result should be viewed with several caveats. The dataset used in this analysis was larger and more varied than any previously analyzed, consisting of many collections. Direct comparisons of machine learning

classification methods are difficult as there are no universally accepted schema to divide collections into training and test sets (unlike standardized competitions in other fields, such as the ImageNet Large Scale Visual Recognition Challenge (ILSVRC)[84]). Furthermore, our exclusion criteria differed, and, because we opted to use multiple scanning sessions from single subjects during training, we also used followup data in ABIDE not employed in previous studies. Class balancing may also have significantly affected the classification accuracy. However, this was necessary to avoid spuriously large accuracies due to the highly skewed ratios of ASD-to-TD individuals. Lastly, preprocessing methods and exclusion criteria are not typically shared across collections, and thus technical and demographic differences in the input data cannot be discounted.

While in this study (and all previous large sample-size studies of ASD classification), the classification percentage of ASD v TD datasets does not approach the standards of clinical diagnosis, but remains pertinent. First, the intention of the models is to encourage further research and analysis in this field. Second, FC data may simply lack discrete, distinguishing signals indicative of ASD, making perfect classification impossible, in which case deep learning ought to be viewed as an advanced statistical model rather than a potential diagnostic tool. Third, ASD is a spectrum and not binary (unlike resting-state/task and, in the vast majority of cases, biological gender), and these labels were applied with varying diagnostic standards. While we are simply using the information available, we recognise that the problem itself may be ill-formed. This is also a potential explanation for the variance in model accuracies seen in Figure 2, compared to the other classification problems addressed. Fourth, due to the influence of confounding factors, high accuracy in machine learning for scientific applications should be viewed with skepticism;[85] for instance, we used several stringent motion-regression algorithms in preprocessing, which likely mitigated the effects of group differences in motion that has previously been observed between autistic and nonautistic subjects.[86]

Finally, our deep learning model provides several advantages and unique features. First, it employed multichannel input. Although this has long been the standard in 2D image classification (for instance, RGB images), it has not been utilized be-

fore in the classification of connectomes. Theoretically, this provides an advantage since it encodes more information about the underlying time-series. In supplementary tests, multichannel inputs generally increased the accuracy of our model by 23% over single-channel Pearson correlation input, though this was not tested extensively. Second, it used vertical filters to encode matrices. In initial versions of this study, we opted to copy the framework of Kawahara et al 2017, which used cross-shaped filters, although this was found to not increase accuracy over vertical filters and caused the model to sometimes fail. Vertical filters were found to be more compatible with the frameworks of modern deep learning libraries, even though they sacrifice the theoretical advantage of encoding edge-to-edge connections.

In our training scheme, we also found substantial accuracy increases with the use of "ensemble" models in machine learning (Table 2); that is, using many independent NNs to vote on a single datapoint. This idea is not new in machine learning,[87–89] but it is notable because the ensemble showed a substantial increase in AUROC and accuracy over the sum of the individual models, and thus in this context it was an effective method of smoothing out unexpected behaviour in models for potential real-world applications. Additionally, it is an effective way to evaluate the performance of a model across the entirety of a dataset. Combined with the attractiveness of evaluating and averaging models independently to reduce variance in class activation, this makes a good case for classifying functional connectomes using many independent models rather than one.

5. Conclusion

Our investigation was the first to amass an exceedingly large and diverse collection of fMRI data and then apply big data methods. We opted to present three important classification tasks and focus on the one that is both most interesting and least-understood. With careful class-balancing, we show that deep learning models are capable of good-quality classifications across mixed collections detecting differences in brain networks, and functions of localized structures, or FCs over large areas. CAMs highlighted key spatial elements of the classification, and our results were largely validated by prior findings of specific phenotypic differences. Activation maximisation gave insights into the types of features

on which the CNN based its classification. While the deep learning model in its present form should not be viewed as a diagnostic tool, it is an example of the apparatus needed to statistically analyse large and publicly accessible volumes of data.

The classification of ASD, on average, pointed overwhelmingly to two key areas (the right caudate nucleus and the right paracentral lobule), which is consistent with many previous studies of ASD. However, it should be noted that the final AUROC was well below the standard for clinical diagnosis, and the variation of model accuracies across our ensemble was very high, especially in relation to the other two categorical classifications. Thus, the areas observed are unlikely to fully characterise ASD. This variation across our very mixed dataset is related to the difficulties of diagnosing ASD in different contexts, and a binary label applied a spectrum disorder may make for an ill-formed machine learning problem.

The most salient future direction of the present work is to focus on one of the classification problems presented and analyse how class activation maps activate differently for different sorts of data. We can also take advantage of several aspects of the technique not explored in the present work, such as comparing the class activations with respect to different input classes. While this was outside the scope of the present study and would have complicated the analysis significantly, it is one of many possible directions in which to take future endeavours.

Bibliography

1. F. Cauda, E. Geda, K. Sacco, F. D'Agata, S. Duca, G. Geminiani and R. Keller, Grey matter abnormality in autism spectrum disorder: an activation likelihood estimation meta-analysis study, *J Neurol Neurosurg Psychiatry* **82** (2011) p. 13041313.

2. T. DeRamus and R. Kana, Anatomical likelihood estimation meta-analysis of grey and white matter anomalies in autism spectrum disorders author links open overlay panel, *NeuroImage: Clinical* **7** (2015) p. 525536.

3. J. Yang and J. Hofmann, Action observation and imitation in autism spectrum disorders: an ale meta-analysis of fmri studies, *Brain Imaging and Behavior* **10** (2015) p. 960969.

4. K. Button, J. Ioannidis, C. Mokrysz, B. Nosek, J. Flint, E. J. Robinson and M. Munafo, Power failure: why small sample size undermines the reliability of neuroscience, *Nat Rev Neurosci* **14** (2013) 365–376.

5. C. Nord, V. Valton, J. Wood and J. Roiser, Power-up: A reanalysis of 'power failure' in neuroscience using mixture modeling, *J Neurosci* **37** (2017) p. 80518061.

6. S. Haar, S. Berman, M. Behrmann and I. Dinstein, Anatomical abnormalities in autism?, *Cereb Cortex* **26** (2016) p. 14401452.

7. W. Zhang, W. Groen, M. Mennes, C. Greven, J. Buitelaar and N. Rommelse, Revisiting subcortical brain volume correlates of autism in the abide dataset: effects of age and sex, *Psychological Medicine* **48** (2018) p. 654668.

8. B. Khundrakpam, J. Lewis, P. Kostopoulos, F. Carbonell and A. Evans, Cortical thickness abnormalities in autism spectrum disorders through late childhood, adolescence, and adulthood: A large-scale mri study, *Cereb Cortex* **27** (2017) p. 17211731.

9. D. van Rooij, E. Anagnostou, C. Arango, G. Auzias, M. Behrmann, G. Busatto, S. Calderoni, E. Daly, C. Deruelle, A. Di Martino *et al.*, Cortical and sub-cortical brain morphometry differences between patients with autism spectrum disorder and healthy individuals across the lifespan: Results from the enigma asd working group, *American Journal of Psychiatry* **175** (2017) p. 359369.

10. E. Müller, A. Schuler and G. Yates, Social challenges and supports from the perspective of individuals with asperger syndrome and other autism spectrum disabilities, *Autism* **12** (2008) p. 173190.

11. T. Simas, S. Chattopadhyay, C. Hagan, P. Kundu, A. Patel, R. Holt, D. Floris, J. Graham, C. Ooi, R. Tait *et al.*, Semi-metric topology of the human connectome: Sensitivity and specificity to autism and major depressive disorder, *PLoS One* **10** (2015).

12. M. Ahmadlou, H. Adeli and A. Adeli, Fractality and a wavelet-chaos-neural network methodology for eeg-based diagnosis of autistic spectrum disorder, *Journal of Clinical Neurophysiology* **27** (2010) p. 328333.

13. M. Ahmadlou, H. Adeli and A. Adeli, Fuzzy synchronization likelihood-wavelet methodology for diagnosis of autism spectrum disorder, *Journal of Neuroscience Methods* **211** (2012) p. 203209.

14. S. Bhat, U. Acharya, H. Adeli, G. Muralidhar Bairy and A. Adeli, Automated diagnosis of autism: In search of a mathematical marker, *Reviews in the Neurosciences* **25** (2014) p. 851861.

15. S. Bhat, U. Acharya, G. Adeli, Muralidhar Bairy, and A. Adeli, Autism: Cause factors, early diagnosis and therapies, *Reviews in the Neurosciences* **25** (2014) p. 841850.

16. A. Antoniades, L. Spyrou, D. Martin-Lopez, A. Valentin, G. Alarcon, S. Sanei and C. Took, Deep neural architectures for mapping scalp to intracranial eeg, *Intl J Neur Sys* **28** (2018).

17. C. Hua, H. Wang, H. Wang, S. Lu, C. Liu, and S. Khalid, A novel method of building functional brain network using deep learning algorithm with application in proficiency detection, *Intl J Neur Sys* **29** (2019) p. 1850015.

18. A. Ansari, P. Cherian, A. Caicedo, G. Naulaers, M. De Vos and S. Van Huffel, Neonatal seizure detection using deep convolutional neural networks, *Intl J Neur Sys* **29** (2019) p. 1850011.

19. F. Schaper, Y. Zhao, M. Janssen, G. Wagner, A. Colon, D. Hilkman, E. Gommer, M. Vlooswijk, G. Hoogland, L. Ackermans *et al.*, Single cell recordings to target the anterior nucleus of the thalamus in deep brain stimulation for patients with refractory epilepsy, *Intl J Neur Sys* **29** (2019).

20. U. Acharya, S. Oh, Y. Hagiwara, J. Tan, H. Adeli and D. Subha, Automated eeg-based screening of depression using deep convolutional neural network, *Comp Meth and Progr in Biomed* **161** (2018) p. 103113.

21. U. Acharya, S. Oh, Y. Hagiwara, J. Tan, and H. Adeli, Deep convolutional neural network for the automated detection of seizure using eeg signals, *Computers in Biology and Medicine* **100** (2018) p. 270278.

22. E. Finn, X. Shen, D. Scheinost, M. Rosenberg, J. Huang, M. Chun, X. Papademetris and R. Constable, Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity, *Nat Neurosci.* **18** (2015) p. 16641671.

23. W. Wang, J. Liu, S. Shi, T. Liu, L. Ma, X. Ma, J. Tian, Q. Gong and M. Wang, Altered resting-state functional activity in patients with autism spectrum disorder: A quantitative meta-analysis, *Front. Neurol.* **9** (2018) p. 556.

24. M. Plitt, K. Barnes and A. Martin, Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards, *NeuroImage Clinical* **7** (2015) p. 35966.

25. M. Just, V. Cherkassky, T. Keller and N. Minshew, Cortical activation and synchronization during sentence comprehension in high-functioning autism: evidence of underconnectivity, *Brain* **127** (2004) p. 181121.

26. V. Cherkassky, R. Kana, T. Keller and M. Just, Functional connectivity in a baseline resting-state network in autism, *Neuroreport* **17** (2006) p. 168790.

27. D. Kennedy and E. Courchesne, The intrinsic functional organization of the brain is altered in autism, *NeuroImage* **39** (2008) p. 187785.

28. M. Assaf, K. Jagannathan, V. Calhoun, L. Miller, M. Stevens, R. Sahl, J. O'Boyle, R. Schultz and G. Pearlson, Abnormal functional connectivity of default mode sub-networks in autism spectrum disorder patients, *NeuroImage* **53** (2010) 247–256.

29. T. Jones, P. Bandettini, L. Kenworthy, L. Case, S. Milleville, A. Martin and R. Birn, Sources of group differences in functional connectivity: an investigation applied to autism spectrum disorder, *NeuroImage* **49** (2010) p. 401414.

30. S. Weng, J. Wiggins, S. Peltier, M. Carrasco, S. Risi, C. Lord and C. Monk, Alterations of resting state functional connectivity in the default network in adolescents with autism spectrum disorders, *Brain Res* **1313** (2010) p. 202214.

31. L. Cerliani, M. Mennes, R. Thomas, A. Martino, M. Thioux and C. Keysers, Increased functional connectivity between subcortical and cortical resting-state networks in autism spectrum disorder, *JAMA Psychiatry* **72** (2015) p. 767777.

32. H. Chien, H. Lin, M. Lai, S. Gau and W. Tseng, Hyperconnectivity of the right posterior temporoparietal junction predicts social difficulties in boys with autism spectrum disorder, *Autism Res* **8** (2015) p. 427441.

33. S. Delmonte, L. O'Gallagher, E. Hanlon, J. McGrath and J. Balsters, Functional and structural connectivity of frontostriatal circuitry in autism spectrum disorder, *Front Hum Neurosci* **7** (2013) p. 430.

34. A. Di Martino, C. Kelly, R. Grzadzinski, X. Zuo, M. Mennes, M. Mairena, C. Lord, F. Castellanos and M. Milham, Aberrant striatal functional connectivity in children with autism, *Biol Psychiatry* **69** (2011) p. 847856.

35. M. Nebel, A. Eloyan, A. Barber and S. Mostofsky, Precentral gyrus functional connectivity signatures of autism, *Front Syst Neurosci* **8** (2014) p. 80.

36. M. Nebel, S. Joel, J. Muschelli, A. Barber, B. Caffo, J. Pekar and S. Mostofsky, Disruption of functional organization within the primary motor cortex in children with autism, *Hum Brain Mapp* **35** (2014) p. 567580.

37. J. Hull, L. Dokovna, Z. Jacokes, C. Torgerson, A. Irimia and J. van Horn, Resting-state functional connectivity in autism spectrum disorders: A review, *Front Psychiatry* **7** (2017).

38. Y. LeCun, P. Haffner, L. Bottou and Y. Bengio, Object recognition with gradient-based learning, *Lecture Notes in Computer Science* **1681** (1999) p. 319345.

39. G. Hinton, S. Osindero and Y.-H. Teh, A fast learning algorithm for deep belief nets, *Neural Computation* **18** (2006) p. 15271554.

40. A. Krizhevsky, I. Sutskever and G. Hinton, Imagenet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems* (2012).

41. J. Bruna, W. Zaremba, , A. Szlam and Y. LeCun, Spectral networks and locally connected networks on graphs, *ICLR* (2014).

42. M. Defferrard, P. Bresson and X. Vandergheynst, Convolutional neural networks on graphs with fast localized spectral filtering, *NIPS* (2016) 3844–3852.

43. W. Hamilton, R. Ying and J. Leskovec, Representation learning on graphs: Methods and applications, *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering* (2017).

44. Y. Hechtlinger, P. Chakravarti and J. Qin, A generalization of convolutional neural networks to graph-structured data, *arXiv* (2017).

45. T. Kipf and M. Welling, Semi-supervised classification with graph convolutional neural networks, *ICLR 2017* (2017).

46. G. Nikolentzos, P. Meladianos, A. Tixier, K. Skianis and M. Vazirgiannis, Kernel graph convolutional neural networks, *ICANN 2018* (2017).

47. M. Jung, H. Kosaka, D. Saito, M. Ishitobi, T. Morita, K. Inohara, M. Asano, S. Arai, T. Munesue, A. Tomoda *et al.*, Default mode network in young male adults with autism spectrum disorder: relationship with autism spectrum traits, *Mol Autism* **5** (2014) p. 35.

48. T. Price, C. Wee, W. Gao and D. Shen, Multiple-network classification of childhood autism using functional connectivity dynamics, *Med Image Comput Comput Assist Interv* **17** (2014) p. 177184.

49. T. Iidaka, Resting state functional magnetic resonance imaging and neural network classified autism and control, *Cortex* **63** (2015) p. 5567.

50. V. Subbaraju, M. Suresh, S. Sundaram and S. Narasimhan, Identifying differences in brain activities and an accurate detection of autism spectrum disorder using resting state functional-magnetic resonance imaging: A spatial filtering approach, *Med Image Anal* **35** (2017) p. 375389.

51. C. Brown, J. Kawahara and G. Hamarneh, Connectome priors in deep neural networks to predict autism, *ISBI 2018* (2018).

52. A. Heinsfeld, A. Franco, R. Craddock, A. Buchweitz and F. Meneguzzia, Identification of autism spectrum disorder using deep learning and the abide dataset, *NeuroImage: Clinical* **17** (2018) p. 1623.

53. M. Khosla, K. Jamison, A. Kuceyeski and M. Sabuncu, 3d convolutional neural networks for classification of functional connectomes, *MICCAI 2018* (2018).

54. K. Simonyan, A. Vedaldi and A. Zisserman, Deep inside convolutional networks: Visualising image classification models and saliency maps, *Workshop at International Conference on Learning Representations*, 2014.

55. J. Kawahara, C. Brown, S. Miller, B. Booth, V. Chau, R. Grunau, J. Zwicker and G. Hamarneh, Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment, *NeuroImage* **146** (2017) 1038–1049.

56. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, *2017 IEEE International Conference on Computer Vision (ICCV)* (2017).

57. D. Erhan, Y. Bengio, A. Courville, and P. Vincent, Visualizing higher-layer features of a deep network, technical report 1341, University of Montreal (2009).

58. R. Poldrack, D. Barch, J. Mitchell, T. Wager, A. Wagner, J. Devlin, C. Cumba, O. Koyejo and M. Milham, Toward open sharing of task-based fmri data: the openfmri project, *Front Neuroinform* **7** (2013).

59. R. Poldrack and K. Gorgolewski, Openfmri: Open sharing of task fmri data, *NeuroImage* **144** (2017) p. 259261.

60. A. Di Martino, C. Yan, Q. Li, E. Denio, F. Castellanos, K. Alaerts, J. Anderson, M. Assaf, S. Bookheimer, M. Dapretto *et al.*, The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism, *Mol Psychiatry* **19** (2014) p. 65967.

61. A. Di Martino, D. O'Connor, B. Chen, K. Alaerts, J. Anderson, M. Assaf, J. Balsters, L. Baxter, A. Beggiato, S. Bernaerts *et al.*, Enhancing studies of the connectome in autism using the autism brain imaging data exchange ii, *Sci Data* **4** (2017) p. 170010.

62. B. Casey and A. Dale, The adolescent brain cogni-

tive development (abcd) study: Imaging acquisition across 21 sites, *Dev Cog Neuro* **32** (2018) p. 4354.

63. D. Hall, M. Huerta, M. McAuliffe and G. Farber, Sharing heterogeneous data: The national database for autism research, *Neuroinformatics* **10** (2012) p. 331339.

64. E. Dolgin, This is your brain online: the functional connectomes project, *Nature Medicine* **16** (2010) p. 351.

65. A. Patel, P. Kundu, M. Rubinov, P. Jones, P. Vertes, K. Ersche, J. Suckling and E. Bullmore, A wavelet method for modeling and despiking motion artifacts from resting-state fmri time series, *NeuroImage* **95** (2014) 287–304.

66. A. Patel and E. Bullmore, A wavelet-based estimator of the degrees of freedom in denoised fmri time series for probabilistic testing of functional connectivity and brain graphs, *NeuroImage* **142** (2016) p. 1426.

67. N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer and M. Joliot, Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain, *NeuroImage* **15** (2002) 273–289.

68. F. Chollet, keras https://github.com/fchollet/keras, (2015).

69. S. Ioffe and C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, *arXiv* (2015).

70. R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, *Intelligence - Volume 2, IJCAI95, Morgan Kaufmann Publishers Inc., San Francisco, C* (1995) 1137–1143.

71. R. Kotikalapudi and contributors, keras-vis https://github.com/raghakot/keras-vis, (2017).

72. A. Qiu, T. Anh, Y. Li, H. Chen, A. Rifkin-Graboi, B. Broekman, K. Kwek, S. Saw, Y. Chong, P. Gluckman *et al.*, Prenatal maternal depression alters amygdala functional connectivity in 6-month-old infants, *Transl Psychiatry* **5** (2015) p. e508.

73. L. Sears, C. Vest, S. Mohamed, J. Bailey, B. Ranson and J. Piven, An mri study of the basal ganglia in autism, *Prog Neuropsychopharmacol Biol Psychiatry* **23** (1999) p. 613624.

74. G. McAlonan, E. Daly, V. Kumari, H. Critchley, T. van Amelsvoort, J. Suckling, A. Simmons, T. Sigmundsson, K. Greenwood, A. Russell *et al.*, Brain anatomy and sensorimotor gating in asperger's syndrome, *Brain* **125** (2002) p. 15941606.

75. P. Brambilla, A. Hardan, S. di Nemi, J. Perez, J. Soares and F. Barale, Brain anatomy and development in autism: review of structural mri studies, *Brain Res Bull* **61** (2003) p. 557569.

76. E. Hollander, E. Anagnostou, W. Chaplin, K. Esposito, M. Haznedar, E. Licalzi, S. Wasserman, L. Soorya and M. Buchsbaum, Striatal volume on magnetic resonance imaging and repetitive behaviors in autism, *Biol Psychiatry* **58** (2005) p. 226232.

77. L. O'Dwyer, C. Tanner, E. van Dongen, C. Greven, J. Bralten, M. Zwiers, B. Franke, D. Heslenfeld, J. Oosterlaan, P. Hoekstra *et al.*, Decreased left caudate volume is associated with increased severity of autistic-like symptoms in a cohort of adhd patients and their unaffected siblings, *PLoS One* **11** (2016) p. e0165620.

78. D. Rojas, E. Peterson, E. Winterrowd, M. Reite, S. Rogers and J. Tregellas, Regional gray matter volumetric changes in autism associated with social and repetitive behavior symptoms, *BMC Psych* **6** (2006) p. 56.

79. K. Turner, L. Frost, D. Linsenbardt, J. McIlroy and R. M uller, Atypically diffuse functional connectivity between caudate nuclei and cerebral cortex in autism, *Behav Brain Funct* **2** (2006).

80. T. Qiu, C. Chang, Y. Li, L. Qian, C. Xiao, T. Xiao, X. Xiao, Y. Xiao, K. Chu *et al.*, Two years changes in the development of caudate nucleus are involved in restricted repetitive behaviors in 25-year-old children with autism spectrum disorder, *Developmental Cognitive Neuroscience* **19** (2016) p. 137143.

81. D. Tomasi and N. Volkow, Gender differences in brain functional connectivity density, *Hum Brain Mapp.* **33** (2013) p. 849860.

82. M. Raichle, A. MacLeod, A. Snyder, W. Powers, D. Gusnard and G. Shulman, A default mode of brain function, *PNAS* **98** (2001) 676–682.

83. M. Greicius, B. Krasnow, A. Reiss and V. Menon, Functional connectivity in the resting brain: A network analysis of the default mode hypothesis, *PNAS* **100** (2003) p. 253258.

84. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla *et al.*, Imagenet large scale visual recognition challenge, *International Journal of Computer Vision* **115** (2015) p. 211252.

85. M. Ribeiro, S. Singh and C. Guestrin, "why should i trust you?" explaining the predictions of any classifier, *Knowledge Discovery and Data Mining (KDD)* (2016).

86. J. Cook, S. Blakemore and C. Press, Atypical basic movement kinematics in autism spectrum conditions, *Brain* **136** (2013) p. 28162824.

87. D. Opitz and R. Maclin, Popular ensemble methods: An empirical study, *Journal of Artificial Intelligence Research* **11** (1999) p. 169198.

88. R. Polikar, Ensemble based systems in decision making, *IEEE Circuits and Systems Magazine* **6** (2006) p. 2145.

89. L. Rokach, Ensemble-based classifiers, *Artificial Intelligence Review* **33** (2010) p. 139.

90. M. Leming and J. Suckling, Deep learning on brain images in autism: What do large samples reveal of its complexity? *Proceedings of the 8th International Work-Conference on the Interplay Between Natural and Artificial Computation, Part I, Springer* 2019, p. 389402.