Analysis of Functional Connectome Pipelines for the Diagnosis of Autism Spectrum Disorders

Clara Jiménez-Valverde[1], Rosa María Maza-Quiroga[1,3], Domingo López-Rodríguez[2], Karl Thurnhofer-Hemsi[1,3(✉)], Ezequiel López-Rubio[1,3], and Rafael Marcos Luque-Baena[1,3]

[1] Department of Computer Languages and Computer Science, University of Málaga, Málaga, Spain
clarajimenez@uma.es, {rosammq,karlkhader,ezeqlr,rmluque}@lcc.uma.es
[2] Department of Applied Mathematics, University of Málaga, Málaga, Spain
dominlopez@uma.es
[3] Instituto de Investigación Biomédica de Málaga - IBIMA, Málaga, Spain

Abstract. This paper explores the effect of using different pipelines to compute connectomes (matrices representing brain connections) and use them to train machine learning models with the goal of diagnosing Autism Spectrum Disorder. Five different pipelines are used to train six different ML models, splitting the data into female, male and all subsets so we can also research the effect of considering male and female patients separately. Our results conclude that pipeline and model choice impact results, along with using general or specific models.

Keywords: Autism · Connectome · Machine learning · Classification

1 Introduction

Autism Spectrum Disorder (ASD) affects the emotional, social, and communication abilities of the patient. Its prevalence among young children is 1–2% [10], but getting the diagnosis is not always easy, and it can require a long process. This is partly due to autism being a spectrum, meaning its characteristics vary between patients and genders. But also because it has been traditionally considered a

This work is partially supported by the Autonomous Government of Andalusia (Spain) under project UMA20-FEDERJA-108, project name Detection, characterization and prognosis value of the non-obstructive coronary disease with deep learning. All of them include funds from the European Regional Development Fund (ERDF). It is also partially supported by the University of Málaga (Spain) under grants B1-2019_01 and B1-2019_02. The authors also thankfully acknowledge the grants of the Universidad de Málaga and the Instituto de Investigación Biomédica de Málaga - IBIMA. Rosa Maza-Quiroga is funded by a Ph.D. grant from the Instituto de Salud Carlos III (ISCIII) of Spain under the i-PFIS program (IFI19/00009).

J. M. Ferrández Vicente et al. (Eds.): IWINAC 2022, LNCS 13259, pp. 213–222, 2022.
https://doi.org/10.1007/978-3-031-06527-9_21

male disease, causing many female patients to be undiagnosed, misdiagnosed, or lately diagnosed [9], which considerably affects their daily lives.

Recently, attention on early diagnosis of ASD through the use of machine learning techniques [3] is increasing. Studies have shown that neurotypical and non-neurotypical brains are wired differently [8], making work on computational modelling of connective differences that can aid in diagnosis important [7].

The study of brain connectivity is based on the construction of the *connectome* [13], a formal representation of the set of brain connections in the form of a weighted graph and its associated adjacency matrix. The graph's nodes generally represent macroscopic regions of the brain, and the weight of its edges indicate the strength of the connection between these regions. This model allows for its management and the application of advanced mathematical and computational techniques. The connectome can be constructed from both anatomical and functional magnetic resonance imaging (fMRI), the latter indicating not the physical connection between brain regions but the intensity of the coactivations between brain regions, i.e., their correlation.

In practice, there are several methods [2,4] to calculate the connectome from an MRI acquisition. This is a problem because (a) there is no standard method for calculation, and (b) the sensitivity of study results to the calculation method used has not been studied. This last point is of great importance, as the possible variability of the results of early diagnostic studies according to the use of different calculation methods is unknown.

This paper aims to analyze the effect of the selection of a particular calculation method on the early diagnosis of ASD using different machine learning methods. In this way, it will be possible to determine whether there is a strong dependence of the results on the *pipeline* of the connectome construction and to discuss how machine learning methods can achieve results comparable to those known in the literature.

The rest of this work is divided into three further sections. We begin by describing the dataset, pipelines, machine learning algorithms and methodology used to create the models. Their results are presented and discussed during the third section, followed by the conclusions.

2 Methodology

The same data has been used on all pipelines to correctly compare their effect on the connectomes produced. While the preprocessed data can be easily obtained for four of the pipelines, the fifth one needs to be executed on the original fMRIs.

Once we have computed the connectomes, all five data sets follow the same process. Firstly, each of them is divided into female, male, and all data sets, which will help us determine the importance of considering each sex separately. For each of these sets (15 in total), we trained six different machine learning (ML) classification models using Cross-Validation and Random Search. A general scheme of the methodology used can be found in Fig. 1.

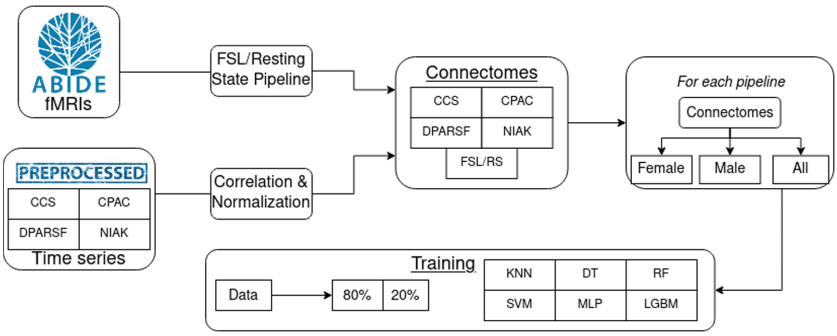

Fig. 1. Scheme of the analysis methodology applied on this study.

2.1 Dataset

The connectomes are computed from the files included in the first edition of the ABIDE project [6], which compiles fMRIs from different entities, totalling 1110 images from both ASD and control patients. The second edition's data has not been used because it is not included on the ABIDE Preprocessed Project [4], which provides time series of four pipelines with several configurations.

- 538 Patients with autism: 65 female and 473 male.
- 572 Patients with autism: 99 female and 473 male.

2.2 Functional Preprocessing Pipelines

A pipeline is a series of steps that, from a raw functional image sequence of a subject, build their correlation matrix (connectome). This work compares five fMRI processing pipelines designed to obtain the connectome. Usually, these pipelines are built from algorithms and functions of standard Neuroimaging processing packages, such as FSL, SPM, ANTs, or FreeSurfer.

The typical steps followed are: S1) basic image pre-processing, correcting artifacts at the beginning or end of the fMRI acquisition, as well as those due to patient motion, slice timing, usually accompanied by a reorientation in a standard coordinate system as well as normalization of the greyscale; S2) removal of signals confusing to process, such as white matter signal, or motion due to the heartbeat and respiration; S3) signal filtering, using, for example, a bandpass filter; S4) transformation (registration) of the image into the standard MNI152 [11] template and labeling to determine a brain parcellation, i.e., the identification of its anatomical regions; S5) calculation of the time series of the average signal (activations) in each anatomical region; S6) construction of the correlation matrix (connectome) between all the time series found in the previous step.

Different pipelines vary on the algorithms used and their parameters. Some add an extra step, usually to eliminate the impact of artifacts in the image.

The first of the pipelines is Duke's Python/FSL Resting State pipeline [2]. It incorporates a non-brain tissue removal stage in step S1), removes the possible

effect of the white matter and cerebrospinal fluid signal in S2), uses a band-pass filter from 0.001 to 0.08 Hz in S3), and, in step S4), uses the Automatic Anatomical Labelling [14] standard for Regions of Interest (ROIs) from Montreal Neurological Institute (aal_MNI) to extract the time series and then the correlation matrices within 116 ROIs. The other pipelines considered in this study are described in the ABIDE website [6] and follow closely the four steps mentioned above, with subtle differences.

The Connectome Computation System (CCS) [15] differs with Duke's pipeline (DUKE) on the software packages used. The Configurable Pipeline for the Analysis of Connectomes (CPAC) [5] removes the effect of the white and grey matter signals, step S2), using principal component analysis instead of linear regression. The Data Processing Assistant for Resting-State fMRI (DPARSF) [16] does not normalize the intensity of the image on step S1). The Neuroimaging Analysis Kit (NIAK) [1] does not correct the timing of each slice, possibly inducing incorrect measurements. On step S2), NIAK removes low-frequency drifts using a discrete cosine basis with a 0.01 Hz high-pass cut-off, whereas the rest of the pipelines apply polynomial regression of the signal.

In the ABIDE platform, the results of the latter four pipelines are pre-computed and downloadable according to several preset settings. In our work, we have collected the data from the platform setting the parameters as close as possible to the ones used in Duke's pipeline since its results are not available online and were computed in-house. The main configurations set are: regarding bandpass filtering, the range is 0.01–0.1 Hz; signal regression is performed using the image's global average on every slice, contrary to the mean white matter and spinal fluid signals, more sophisticated; and while Duke's pipeline uses the aal_MNI atlas for ROIs, they use the aal atlas, with the same labels for ROIs, but a different template and a slightly different coordinate system.

For all pipelines, the resulting time series is used to compute the normalized connectomes, as the normalized correlation matrices between them.

2.3 Machine Learning Methods

The six supervised machine learning classifiers used are K-nearest neighbors, Decision Trees, Random Forest, Support Vector Machines, Multilayer Perceptron, and LightGBM, using the algorithms available on scikit-learn's Python package [12]. Training has been performed separately for each pipeline and dataset (female, male, and both), and 5-fold cross-validation combined with Random Search for each chunk has been used to ensure the models' validity. The search grid for each model has been adjusted individually for each pipeline and dataset to aim at the best possible results.

- **K-Nearest-neighbors** (KNN) is one of the most basic classifiers since it simply stores training data and their class and then compares new data to its k nearest neighbors. The class with the most neighbors is assigned.
- **Decision Trees** (DT) creates models based on rules with higher complexity as the tree's depth increases. These rules are created by dividing training data

into two iteratively, basing divisions on differences between each subset and applying them when certain information gain is achieved.
- **Random Forests** (RF) is an ensemble method, meaning it combines different models to obtain better results. It uses a series of Decision Trees' probabilistic results to reach a verdict.
- **Support Vector Machines** (SVM) create a series of hyper-planes to split training data into its different classes. The best hyperplanes are chosen based on their distance to the nearest points.
- **Multilayer Perceptron** (MLP) belongs to the Neural Networks family. It contains an input and an output layer, with at least one hidden layer in between containing the main weights of the model.
- **LightGBM** (LGBM) uses decision tree-based algorithms seeking better accuracy with higher speed and efficiency while using less memory and therefore handling large-scale data. It achieves this with histogram-based algorithms and best-first growing criteria.

3 Experimental Results

This section presents the experimental results obtained using the six ML methods specified above. The five pipelines analyzed in this paper were compared for each method and each subset of the data (male, female, and all data).

For this purpose, the mean accuracy, sensitivity (True Positive Rate, TPR), and specificity (True Negative Rate, TNR) were computed for train and test data using 5-fold CV combined with random search:

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

$$TPR = \frac{TP}{TP + FN}, \qquad TNR = \frac{TN}{TN + FP} \tag{2}$$

where TP, TN, FP and FN are the true positives, true negatives, false positives, and false negatives, respectively.

First, the results for female data are presented in Figs. 2, 3, and Table 1. KNN and DT models produced the worst trainings, except for the DUKE pipeline using KNN. However, it is probably due to overfitting since the test accuracy is not remarkable. On the other hand, the best results on the training sets were obtained using the MLP and LGBM models, reaching almost 100% accuracy. Now, if we focus on the test sets (Fig. 3), i.e., unseen data, we can have a better overview of the performance of each method. It is shown that SVM and MLP are the best options, specifically for CCS, CPAC, and DPARSF pipelines, reaching accuracies around 70%.

Although CCS and CPAC are the most stable protocols, there is no clear optimum pipeline, meaning that the type of machine learning method used does not matter since they provide the best classification metric for female data. DUKE and NIAK protocols are not recommendable for any method.

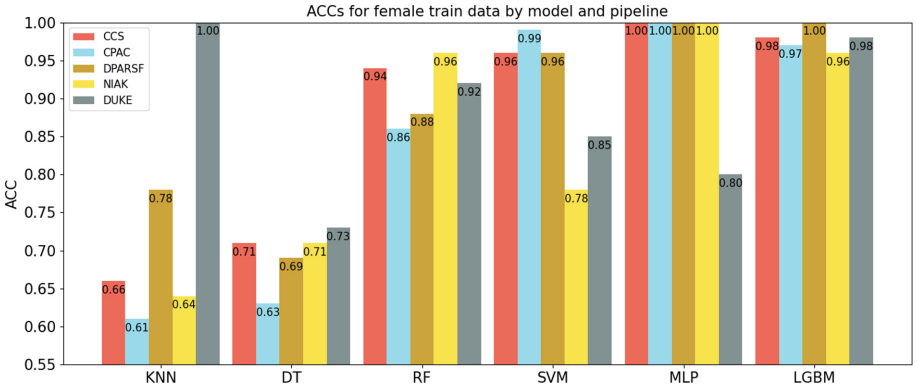

Fig. 2. Mean ACCs for female train data using 5-fold cross-validation.

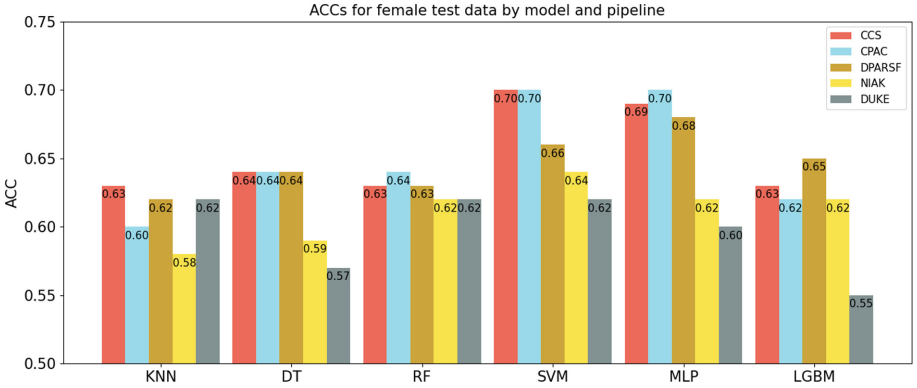

Fig. 3. Mean ACCs for female test data using 5-fold cross-validation.

Table 1. Average Sensitivity (TPR) and Specificity (TNR) measures for female data on the test set using 5-fold cross-validation. Best results are in bold.

	CCS		CPAC		DPARSF		NIAK		DUKE	
	TPR	TNR	TPR	TNR	TPR	TNR	TPR	TNR	TPR	TNR
KNN	0.63	0.72	0.61	0.20	0.62	0.53	0.60	0.07	0.62	0.67
DT	0.66	0.55	0.64	0.16	0.65	0.64	0.64	0.27	0.61	0.26
RF	0.63	0.60	0.63	0.68	0.63	0.51	0.63	0.71	0.62	0.57
SVM	0.70	0.73	**0.72**	0.71	0.66	0.69	0.64	0.70	0.64	0.61
MLP	0.71	0.64	0.70	**0.76**	0.68	0.67	0.65	0.54	0.64	0.51
LGBM	0.65	0.56	0.65	0.54	0.66	0.62	0.65	0.52	0.60	0.40

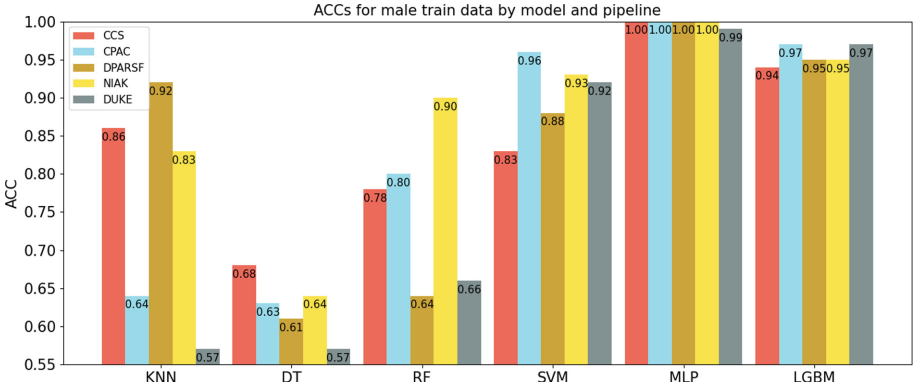

Fig. 4. Mean ACCs for male train data using 5-fold cross-validation.

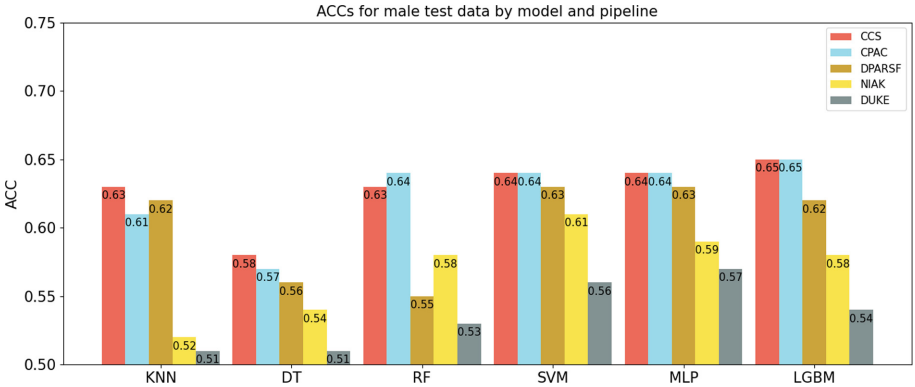

Fig. 5. Mean ACCs for male test data using 5-fold cross-validation.

Table 1 summarizes the TPRs and TNRs obtained by each model and pipeline. As in other medical diagnosis tasks, having a low false negative rate is essential. Thus, a high TNR is an objective. The best measures were achieved using the CPAC pipeline, reaching 72% TPR and 76% TNR for SVM and MLP models, respectively. Specifically, MLP seems to be the most accurate for both measures. CCS pipeline also gives good rates using the SVM model. The rest of the pipelines remain behind, being the KNN model the worst classifier.

Moving on to results for male data (shown in Figs. 4, 5, and Table 2), the overall best training results were obtained once again with MLP and LGBM, while DT gave the worst ones by far. Looking at the test sets, accuracies do not vary too much between models, excepting DT, which has the lowest overall scores. The used pipeline has influence, with CCS and CPAC giving the best results (up to 65% accuracy) and DUKE the worst.

The sensitivities and specificities have decreased a bit concerning female data, reaching a maximum of 65% TPR and 66% TNR. Nevertheless, the larger

Table 2. Average Sensitivity (TPR) and Specificity (TNR) measures for male data on the test set using 5-fold cross-validation. Best results are in bold.

	CCS		CPAC		DPARSF		NIAK		DUKE	
	TPR	TNR	TPR	TNR	TPR	TNR	TPR	TNR	TPR	TNR
KNN	0.63	0.63	0.60	0.63	0.62	0.64	0.52	0.52	0.53	0.54
DT	0.59	0.59	0.58	0.57	0.57	0.56	0.53	0.55	0.50	0.42
RF	0.64	0.62	0.64	0.63	0.59	0.54	0.59	0.57	0.53	0.54
SVM	0.64	0.64	0.64	0.64	0.63	0.64	0.60	0.62	0.56	0.57
MLP	0.64	0.64	0.64	0.65	0.63	0.64	0.58	0.61	0.57	0.57
LGBM	**0.65**	0.65	0.64	**0.66**	0.62	0.63	0.58	0.59	0.54	0.55

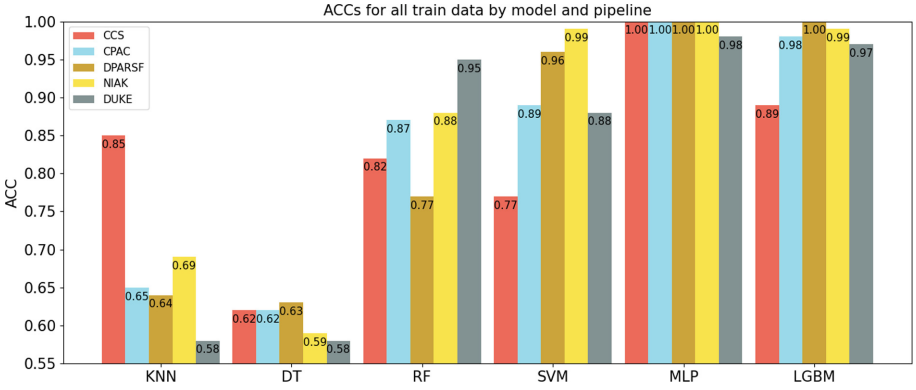

Fig. 6. Mean ACCs for all train data using 5-fold cross-validation.

amount of data (around five times more than female) has provoked a better classification for KNN and DT, giving now reasonable results. Still, CCS and CPAC provide the best rates while DUKE almost does not overcome the 50%.

Results for the subset containing all data are found in 6, 7, and Table 3. Training accuracies show better results for MLP and LGBM and worse ones for DT, once more. MLP offers the best results, reaching 70% accuracy on CPAC preprocessing for the test set. SVM is also close to that score. CCS is again the second-best pipeline, and NIAK and DUKE have inferior performance.

Table 3 shows that CCS pipeline are highly competitive, having the best TNR (68%). This is a more reliable value since we have a big and diverse dataset. MLP and SVM are again the best classifiers, depicting similar sensitivity and specificity (67%). It is also remarkable that a simpler model such as KNN performs better or at least has the same TNR. Clearly, DUKE is the worst pipeline.

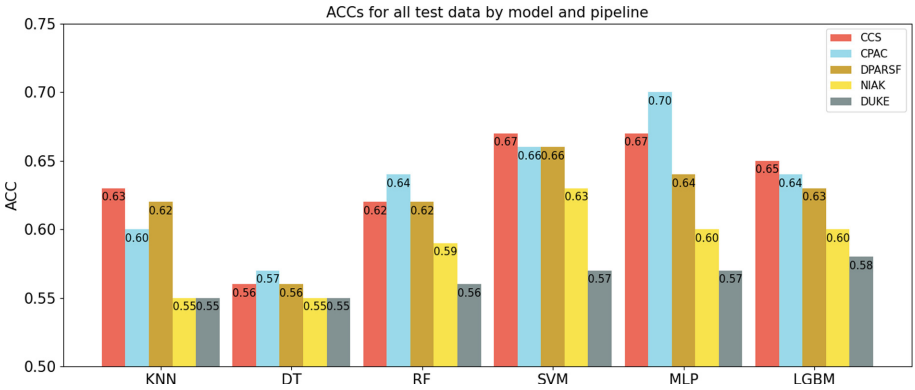

Fig. 7. Mean ACCs for all test data using 5-fold cross-validation.

Table 3. Average Sensitivity (TPR) and Specificity (TNR) measures for all data on the test set using 5-fold cross-validation. Best results are in bold.

	CCS		CPAC		DPARSF		NIAK		DUKE	
	TPR	TNR	TPR	TNR	TPR	TNR	TPR	TNR	TPR	TNR
KNN	0.62	**0.68**	0.58	0.67	0.60	0.67	0.55	0.55	0.57	0.53
DT	0.57	0.57	0.58	0.57	0.58	0.56	0.56	0.43	0.55	0.57
RF	0.61	0.64	0.64	0.66	0.61	0.65	0.58	0.61	0.57	0.55
SVM	0.66	**0.68**	**0.67**	0.66	**0.67**	0.66	0.63	0.63	0.57	0.57
MLP	**0.67**	**0.68**	**0.67**	0.66	0.65	0.63	0.60	0.60	0.58	0.56
LGBM	0.65	0.65	0.64	0.64	0.64	0.62	0.60	0.60	0.59	0.57

4 Conclusions

After analyzing our results, we can extract a series of conclusions. Generally, their accuracies are not great, since the maximum values achieved are 70% for female and all data and 65% for male data. The best TPRs and TNRs are also close to these values, the best combination being 70% TPR and 76% TNR for the MLP model using female CPAC-processed data. It seems like female-oriented models are more specific and sensitive than general models, but when it comes to male patients, using male-specific models worsens the results.

Regarding pipeline choice, it does impact results, with CCS and CPAC pipelines being the most reliable ones, constantly providing results among the best. NIAK and DUKE are the opposite, usually resulting on lower values. The one subset where results are more balanced between pipelines is the female one, with less harsh differences. Model choice also impacts the outcome. SVM and MLP models repeatedly stand out as the ones with better results, with the addition of LGBM on male data and KNN on all data. DT, on the other hand, is consistently the worst one.

Future works could explore how results vary using different configurations on the pipelines, or changing the atlas used to extract ROIs. Other machine learning methods could also be researched and the application of these same methods on connectomes to diagnose other disorders such as schizophrenia or depression.

References

1. Bellec, P., et al.: A neuroimaging analyses kit for Matlab and octave. In: Human Brain Mapping HBM 2011 17th Annual Meeting of the Organization on Human Brain Mapping, Quebec City, Canada, 26–30 June 2011, pp. 1–5. Organization on Human Brain Mapping (2011)
2. BIAC resting state pipeline software [Brain Imaging & Analysis Center]. https://wiki.biac.duke.edu/biac:analysis:resting_pipeline
3. Bone, D., et al.: Applying machine learning to facilitate autism diagnostics: pitfalls and promises. J. Autism Dev. Disord. **45**(5), 1121–1136 (2015)
4. Craddock, C.A., et al.: The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. Front. Neuroinformat. **7** (2013)
5. Craddock, C., et al.: Towards automated analysis of connectomes: the configurable pipeline for the analysis of connectomes (C-PAC). Front. Neuroinform. **42**, 10–3389 (2013)
6. Martino, D., et al.: The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol. Psychiatry **19**(6), 659–667 (2014)
7. Heinsfeld, A., et al.: Identification of autism spectrum disorder using deep learning and the abide dataset. NeuroImage: Clinical **17**, 16–23 (2018)
8. Hong, S.J., et al.: Atypical functional connectome hierarchy in autism. Nat. Commun. **10**(1), 1–13 (2019)
9. Lockwood Estrin, G., et al.: Barriers to autism spectrum disorder diagnosis for young women and girls: a systematic review. Rev. J. Autism Dev. Disorders (2020)
10. Maenner, M.J., et al.: Prevalence of autism spectrum disorder among children aged 8 Years-Autism and developmental disabilities monitoring network, 11 Sites, United States, 2016. MMWR Surveill. Summ. **69**(4), 1–12 (2020)
11. Mazziotta, J.C., et al.: A probabilistic atlas of the human brain: theory and rationale for its development. Neuroimage **2**(2), 89–101 (1995)
12. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
13. Sporns, O., et al.: The human connectome: a structural description of the human brain. PLoS Comput. Biol. **1**(4), e42 (2005)
14. Tzourio-Mazoyer, N., et al.: Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage **15**(1), 273–289 (2002)
15. Xu, T., et al.: A connectome computation system for discovery science of brain. Sci. Bull. **60**(1), 86–95 (2015)
16. Yan, C., Zang, Y.: Dparsf: a matlab toolbox for "pipeline" data analysis of resting-state fmri. Front. Syst. Neurosci. **4**, 13 (2010)