7-2019

Auto-ASD-Network: A technique based on Deep Learning and Support Vector Machines for diagnosing Autism Spectrum Disorder using fMRI data

Taban Eslami

Fahad Saeed

Auto-ASD-Network: A technique based on Deep Learning and Support Vector Machines for diagnosing Autism Spectrum Disorder using fMRI data

TABAN ESLAMI, Department of Computer Science
Western Michigan University
FAHAD SAEED*, School of Computing and Information Science
Florida International University, Florida

Quantitative analysis of brain disorders such as Autism Spectrum Disorder (ASD) is an ongoing field of research. Machine learning and deep learning techniques have been playing an important role in automating the diagnosis of brain disorders by extracting discriminative features from the brain data. In this study, we propose a model called Auto-ASD-Network in order to classify subjects with Autism disorder from healthy subjects using only fMRI data. Our model consists of a multilayer perceptron (MLP) with two hidden layers. We use an algorithm called SMOTE for performing data augmentation in order to generate artificial data and avoid overfitting, which helps increase the classification accuracy. We further investigate the discriminative power of features extracted using MLP by feeding them to an SVM classifier. In order to optimize the hyperparameters of SVM, we use a technique called Auto Tune Models (ATM) which searches over the hyperparameter space to find the best values of SVM hyperparameters. Our model achieves more than 70% classification accuracy for 4 fMRI datasets with the highest accuracy of 80%. It improves the performance of SVM by 26%, the stand-alone MLP by 16% and the state of the art method in ASD classification by 14%.

The implemented code will be available as GPL license on GitHub portal of our lab (https://github.com/PCDS).

Additional Key Words and Phrases: fMRI; Time series; Pearson's Correlation Coefficient; ASD; Deep Learning; Classification; MLP

1 INTRODUCTION

Diagnosing brain disorders such as Alzheimer's, Mild Cognitive Impairment (MCI), Attention Deficit Hyper Activity Disorder (ADHD) and Autism Spectrum Disorder (ASD) using machine learning and deep learning techniques is an ongoing field of research [7, 11, 15, 17, 19, 29]. In this study, we focus on ASD disorder which is diagnosed in more than 1% of children. ASD is a neurological and developmental brain disorder which affects the social communication and behaviour of the children. This disorder is not curable and continues to adulthood. ASD diagnosis currently relies on traditional methods like screening the child's behaviour and interviewing parents [25]. These methods are error prone which may cause harmful side effects due to overprescribing drugs [27]. In order to diagnose brain disorders like ASD in a more quantitative manner, research has been pushed towards analyzing brain imaging data such as Magnetic Resonance Imaging (MRI) and functional Magnetic Resonance Imaging (fMRI) using machine learning and deep learning techniques.

Generally, MRI and fMRI techniques, provide images from different levels of the brain which are widely used as the input of machine learning techniques. In fMRI data, the brain is divided into small cubic elements called voxels and the activity of each voxel over time is extracted as a time series. The statistical association between two voxels which is also known as their functional connectivity is defined as the correlation between their time series values. Pearson's

*Corresponding author

Authors' addresses: Taban Eslami, Department of Computer Science
Western Michigan University, Kalamazoo, MI, , taban.eslami@wmich.edu; Fahad Saeed, School of Computing and Information Science
Florida International University, Miami, Florida, fsaeed@fiu.edu.

correlation is the most widely used measure for computing functional connectivity and is calculated using the following equation:

$$\rho_{uv} = \frac{\sum_{t=1}^{T}(u_t - \bar{u})(v_t - \bar{v})}{\sqrt{\sum_{t=1}^{T}(u_t - \bar{u})^2}\sqrt{\sum_{t=1}^{T}(v_t - \bar{v})^2}} \tag{1}$$

In which u and v are two time series of length T and $\bar{u}$ and $\bar{v}$ correspond to their mean value respectively.

Alteration in functional connectivities may cause different brain disorders such as Alzheimer's, Schizophrenia, and other disorders [21, 24] which motivated us to consider functional connectivities as the features of our model.

Autism Brain Imaging Data Exchange (ABIDE) initiative has provided a dataset containing fMRI and MRI data generated from 1112 healthy control and ASD subjects. The data is coming from 17 different brain imaging centers. Researchers have come up with new techniques for diagnosing ASD using MRI [18] and fMRI [1, 6, 14, 16] data provided by ABIDE repository. Some of these techniques are based on conventional machine learning techniques such as Support Vector Machines (SVM) and Random Forest [1, 3, 12]. For example, Bi et al [3], used the random SVM cluster for classification of healthy from ASD subjects. Various studies used the demographic information of the subjects such as age, IQ, and handness in their methods or selected subsets of subjects with specific attributes in their analysis. For instance, Parisot et al. [26], represented the population of the subjects as a graph in which imaging features correspond to vertices and phenotypic information of the subjects define the weights of the edges connecting them to each other. There are few studies such as [14] which used only fMRI data without considering any demographic information in their analysis. Using only fMRI data for classifying ASD vs healthy subjects provides a tool for clinicians to assist them in decision making process without being biased with other demographic information. Although including other information may increase the prediction accuracy, our goal is to rely solely on brain fMRI data for detecting ASD disorder, which is a more challenging task.

Deep Learning techniques such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) and Autoencoders have gained astonishing success in the past few years in many different areas including computer vision, speech recognition, health care and etc. Different deep learning based methods are also proposed for diagnosing ASD [2, 4, 10, 13, 20, 23]. For example, Devornel et al. [10] used Long Short-Term Memory Networks (LSTMs) and obtained 68.5% classification accuracy. Heinsfeld et al. [14] used two denoising autoencoders in order to extract a lower dimensional feature vector from the data and used their weights as the initial weights of an MLP. Fine-tuning the MLP resulted in 70% classification accuracy. To the best of our knowledge, this method is the state of the art technique in ASD classification of whole ABIDE dataset.

Deep Learning models contain a huge number of parameters that should be optimized during the learning process. Providing large training data for deep learning models, makes them more general to unseen data and avoids problems like overfitting. Unfortunately, large datasets are not always available. This is the case in the field of neuroimaging in which scanning subjects and generating more data is a time consuming and costly task. In such cases, techniques like data augmentation can be useful. Data augmentation methods are shown to be helpful in reducing overfitting and generalizing deep learning models [22].

In this paper, we propose a deep learning based approach for classification of healthy subjects from ASD patients by *only* using fMRI data and without considering any demographic information. In order to increase the number samples and avoid overfitting, we augment artificial data into training set using a technique called Synthetic Minority Over-Sampling (SMOTE) [5]. SMOTE was originally proposed in order to oversample minority class in imbalanced datasets. In this study, we use SMOTE in order to oversample both healthy and ASD class to double the size of the training set, although the datasets we use here are almost balanced in terms of ASD and healthy subjects. We also investigate the effectiveness of the features extracted using the deep learning model. To do this, besides doing experiments on the deep learning model as a classifier, we use its hidden layer which contains extracted features, as the input to SVM classifier. Considering that SVM has some hyperparameters like kernel function and penalty which their optimal values are not known beforehand, we use a method called Auto Tune Models (ATM) to automate the hyperparameter tuning process. Our experiments show significant improvement in classification accuracy by using the oversampling technique and using SVM.

2 DEEP LEARNING BASED MODEL FOR ASD CLASSIFICATION

The model that is used in this study is a multilayer perceptron with two hidden layers (Fig 2 part D). The input layer of the network receives the pairwise Pearson's correlation coefficients of all regions in the brain (functional connectives) computed using equation 1. Since Pearson's correlation have a symmetric property, instead of using N^2 pairwise correlations, we use $\frac{N(N-1)}{2}$ distinct correlations extracted from upper triangle (or lower triangle) part of the correlation matrix and avoid redundant values. Here N refers to the number of brain regions.

Assuming x_i is the input of layer i, W_i is the vector of weights connecting the nodes in layer i to to the nodes in layer $i+1$ and b_i as bias value of that layer, layer $i+1$ is activated using the following equation:

$$z_{i+1} = f(W_i x_i + b_i) \tag{2}$$

In which f is the rectifier activation function (ReLU) defined as:

$$f(x) = x^+ = \max(0, x) \tag{3}$$

Softmax function is finally applied to the final output layer, which determines the probability that the input feature vector corresponds to each classes. Softmax function is computed using the following equation.

$$z_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}} \tag{4}$$

In which z_i corresponds to the activation of node i. Since in our problem there are two classes (associated with ASD and Healthy groups), the value of k in equation 4 is equal to 2. The node with the highest probability determines the class. The classification loss is then computed using negative log likelihood function as $L_i = -log(p_i)$ In which p_i is the probability of correct class computed using the softmax function. The value of the loss is used to optimize the parameters of the network using backpropagation algorithm.

2.1 Oversampling Using SMOTE

As mentioned earlier, deep learning methods need a large amount of data in order to generate reliable results. One issue we are facing in our classification problem is the lack of enough data, such that each dataset contains less than 200 samples. This motivated us to generate synthetic samples using available data in order to improve the quality of the model and avoid overfitting. There are some traditional data augmentation methods in the field of computer vision such as rotating and flipping images, to generate extra training data, but these techniques will not work in our problem since the nature of features we are using are completely different than images. Instead, we use a method called SMOTE which was originally proposed for oversampling the minority class in imbalanced datasets [5]. In this method, the oversampling is performed by generating one or more synthetic samples per training point in the minority class. For each sample in the minority class, an artificial sample is generated by linear interpolation between two points randomly selected from its k nearest neighbors. This process is repeated for each sample based on the desired number of artificial data. In our proposed approach, we utilize this method for generating one artificial point per existing samples in the training set. Since the length of our features is quite long, we skipped finding k nearest neighbors and instead, picked candidate samples randomly from the same class as the original sample. This process results in doubling the size of the training set and improving the prediction performance which will be discussed more in the experiments section.

2.2 Hyperparameter Tuning using ATM

Hidden layers in a deep neural network are known for learning complex features from the input data. The final layer of the network receives these extracted features from previous layers and performs the classification. We hypothesized that using these features combining with other classification methods such as SVM could result in high accuracy since SVM is known as a very effective method in fMRI classification. In this way, we combine the power of deep learning for extracting features and benefit from the well known SVM method. SVM classifier has different set of hyperparameters, for example the penalty parameter, kernel function and parameters related to specific kernels such as the degree of polynomial in the polynomial kernel function. In order to find the optimal hyperparameters for SVM, we used a tool called Auto-Tuned Models (ATM) [28]. ATM is a scalable multi-method and self optimizing tool which automates and optimizes the hyperparameter tuning process of machine learning algorithms.

ATM implements a parameter-search algorithm by partitioning the hyperparameter space using a conditional-parameter tree (CPT). Each branch of the CPT corresponds to fixed hyperpartitions. An example of CPT for SVM classifier is shown in figure 1. After building the CPT, ATM performs parameter search in two steps. In the first step, the hyperpartitions are selected based on multi-armed bandit. Then, the hyperparameters within each partition are tuned based on Gaussian Process technique. ATM provides different options for tuning hyperpartitions (uniform, multi-armed bandit, hierarchical multi-armed bandit, etc.) and hyperparameters (uniform, Gaussian Process, etc.)

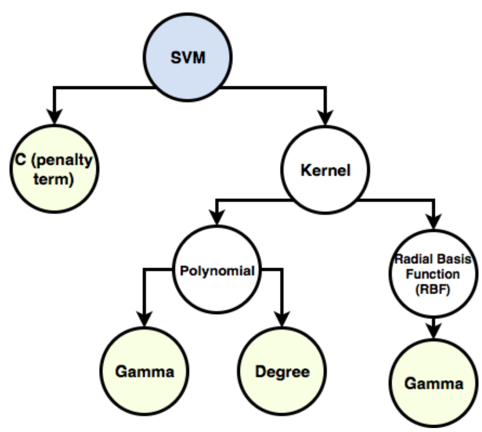

Fig. 1. Example of CPT for SVM classifier

ATM provides a parameter called budget for resource allocation. It could be defined as either the total computation time, or the total number of classifiers to try. For this study, we set the budget to 50 classifiers. After hyperparameter tuning by ATM is finished, we use the top 10 classifiers with the highest Cross-Validation accuracy on training set as the candidate classifiers, and perform the ensemble classification with a voting mechanism to predict the label of each test sample.

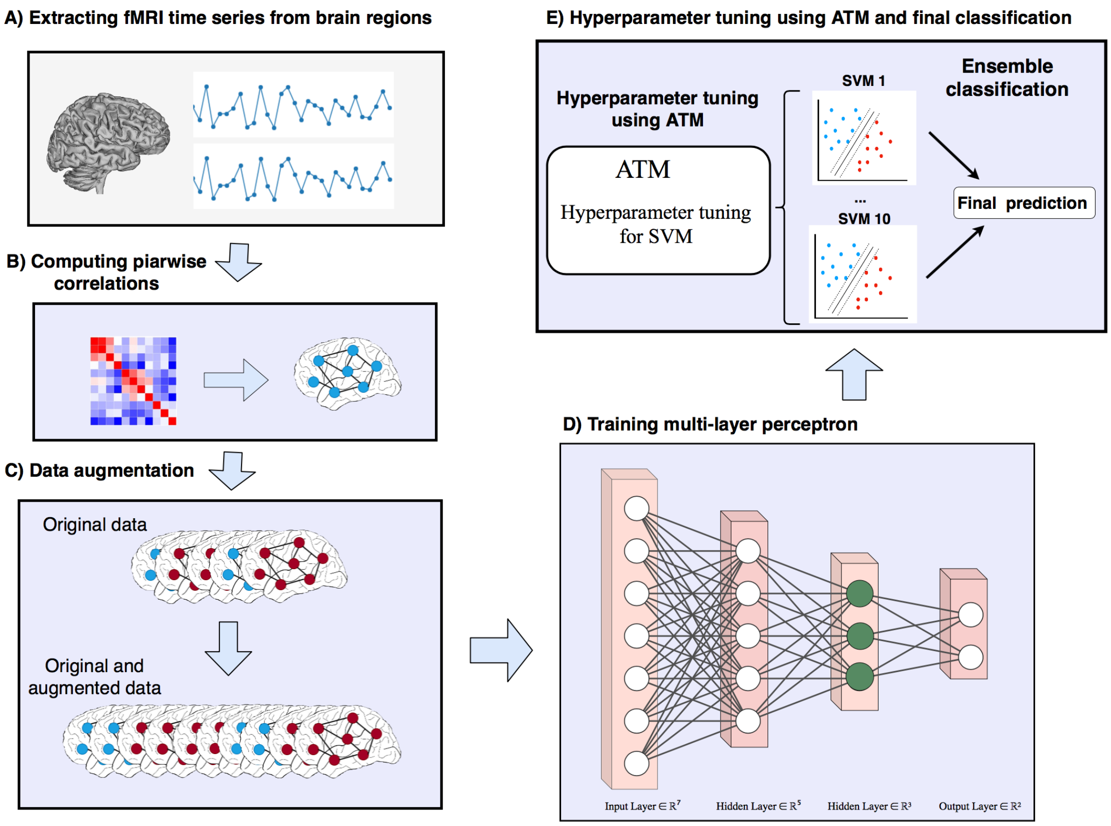

Fig. 2. Overall classification framework of Auto-ASD-Network: A) Time series are extracted from different regions. B) Pairwise functional connections are computed using Pearson's correlation. C) Artificial data is generated in feature space by applying SMOTE algorithm on training data D) multilayer perceptron is trained using training data (This Model can be used for final classification or features can be extracted from the hidden layer and sent to part E. E) ATM is used for finding the best parameters of SVM on features extracted from MLP in Part D. Top 10 classifiers are used for predicting the class label of a hold-out test sample (ensemble classification).

3 EXPERIMENTS AND RESULTS

3.1 ABIDE dataset

As mentioned earlier, ABIDE initiative has gathered and preprocessed brain imaging data from ASD as well as healthy subjects from different brain imaging centers [8]. While different pipelines are used for preprocessing the data, in this study, we used C-PAC pipeline in which preprocessing steps include motion correction, slice timing correction, nuisance signal removal, low frequency drifts, and voxel intensity normalization. ABIDE also provided seven different parcellation methods in which the brain is parcellated to several different regions. The data that we used for this study is parcellated to 200 regions using spatially constrained spectral clustering algorithm [9]. It is worth mentioning that each

Table 1. Class membership information of ABIDE-I fMRI dataset for each individual site

Site	NYU	OHSU	USM	UCLA
ASD	75	12	46	54
Healthy control	100	14	25	44

data center generated brain imaging data using different parameters and scanning protocols. Parameters like repetition time (TR), echo time (TE), and openness or closeness of the eyes during the scan are different among each data center. We used 4 datasets from ABIDE-I repository for conducting our experiments. The class membership information of the datasets are shown in table 1.

3.2 Classification performance

In order to measure the classification performance of our proposed method, we performed 5-fold Cross-validation and compared the average accuracy, sensitivity and specificity of each method. Considering **TN** as the number of correctly classified healthy subjects, **TP** as correctly diagnosed ASD subjects, **FP** as falsely diagnosed with ASD and **FN** as ASD subjects diagnosed as healthy, accuracy of classifier is defined as $\frac{TP + TN}{TP + TN + FP + FN}$, sensitivity as $\frac{TP}{TP + FN}$ and specificity as $\frac{TN}{TN + FP}$.

3.3 List of methods

For all methods listed in this section, the feature vector of each sample is the set of pearson's correlation coefficients between time series of each pair of brain regions. Since the fMRI data that we use is parcellated to 200 regions, each feature vector contains 19900 ($\frac{200 \times 199}{2}$) distinct pairwise correlation coefficients. The methods that are evaluated for ASD classification are as follows:

- Method 1: Ref. [14]
  In this method, first, two denoising autoencoders extract a lower dimensional representation from the input data. Then, the weights of the autoencoders are used as initial weights of an MLP. This MLP is trained on the input data and is used for the final classification.
- Method 2: SVM
  SVM is used as the first baseline classifier which is trained on the original input data (19900 pairwise correlation coefficients).
- Method 3: SVM-ATM
  SVM is used as the classifier which is trained on the original input data. SVM hyperparameters are tuned using ATM technique.
- Method 4: MLP
  MLP is used as the second base classifier which is trained on original input data.
- Method 5: MLP-DA
  MLP is used as the second baseline classifier which is trained on original data as well as artificial data generated using SMOTE algorithm.
- Method 6: MLP-SVM-ATM
  Similar to methods 1 and 2, however, the input data to SVM are features that are extracted from the hidden

layer of MLP. In this case, MLP is trained by using only original data and no data augmentation is performed. Parameters of SVM are optimized using ATM.

- Method 7: Auto-ASD-Network
  Similar to method 6, SVM is used as the final classifier which receives its input from the last hidden layer of the MLP, with the addition of the data augmentation using SMOTE for training the MLP. Parameters of SVM are optimized using ATM.

All the experiments reported in this section are performed on a Linux system containing two Intel Xeon E5-2620 Processors at 2.4 GHz and total 48 GBs of RAM. The system contains an NVIDIA Tesla K-40c GPU with 2880 CUDA cores and 12 GBs of RAM. CUDA version 8 and PyTorch library were used for performing the experiments.

3.4 Evaluating the effect of ATM

In the first experiment, we evaluated the effect of hyperparameter tuning using ATM. Table 2 shows the classification performance of SVM with and without hyperparameter tuning (SVM and SVM-ATM). According to the results in Table 2,

Table 2. Performance comparison of traditional SVM and SVM optimized using ATM (SVM-ATM)

Site	Method	Accuracy	Sensitivity	Specificity
OHSU	SVM	54	0	100
	SVM-ATM	**72.3**	56.6	83.3
NYU	SVM	57.1	0	100
	SVM-ATM	**69.1**	53.3	81
USM	SVM	64.7	100	0
	SVM-ATM	**69.6**	84.3	42
UCLA	SVM	55.1	100	0
	SVM-ATM	**72.2**	83.8	57

ATM significantly improves the performance of SVM classifier. Without tuning hyperparameters, for all datasets either sensitivity or specificity is equal to zero, which means that all test subjects are classified either as healthy or ASD.

3.5 Evaluating the effect of data augmentation

In this experiment, we examined the performance of deep neural network as a classifier, with and without performing data augmentation (i.e. MLP and MLP-DA). We also measured the performance of hyperparameter tuned SVM trained using the features extracted from MLP (i.e. MLP-SVM-ATM and Auto-ASD-Network). The results are shown in Table 3.

As the results imply, data augmentation improves the performance of MLP classifier by increasing classification accuracy. Data augmentation also helps the network to provide better features for SVM classifier, as the performance of Auto-ASD-Network is better than MLP-SVM-ATM. Overall, among the 4 datasets that we used, Auto-ASD-Network outperforms other methods as it shows almost equal or higher accuracy. For OHSU dataset, Auto-ASD-Network significantly outperforms all other methods and achieves 80% accuracy.

Table 3. Performance comparison of different methods with and without data augmentation

Site	Method	Accuracy	Sensitivity	Specificity
OHSU	Ref. [14]	74	66.6	86.6
	MLP	64	62.5	61.6
	MLP-DA	74.3	74.1	70.8
	MLP-SVM-ATM	78	67.3	84.6
	Auto-ASD-Network	**80**	73	83
NYU	Ref. [14]	64.5	78	46
	MLP	68.5	44	87
	MLP-DA	**70**	65.1	71.5
	MLP-SVM-ATM	69.7	57.3	79
	Auto-ASD-Network	**70**	57.9	79.2
USM	Ref. [14]	62	20	84
	MLP	64	100	0
	MLP-DA	70	70	53.7
	MLP-SVM-ATM	72.3	85	42
	Auto-ASD-Network	**72.4**	87.3	45
UCLA	Ref. [14]	57.7	58	57.4
	MLP	71.9	76.7	64.8
	MLP-DA	**72.7**	77.6	65.2
	MLP-SVM-ATM	70.6	75.6	63.6
	Auto-ASD-Network	72.2	82.3	59.8

Table 4. Running time of each method

Method	Running time
Ref [14]	7 min
SVM	4 sec
SVM-ATM	11.25 min
MLP	2.5 min
MLP-DA	4.8 min
MLP-SVM-ATM	5.3 min
Auto-ASD-Network	7.6 min

3.6 Running Time

We measured the running time of different methods on UCLA dataset which is shown in Table 4. Among all methods, SVM with no hyperparameter tuning has the fastest running time but the worst performance. Using ATM for optimizing SVM hyperparameters trained on original feature vectors (19900 pairwise correlations) is the most time-consuming method since ATM needs to train the SVM several times with a large number of features. Data augmentation almost doubles the running time due to the increasing number of training samples.

4 CONCLUSION AND FUTURE WORK

In this paper, we focus on the classification of Autism Spectrum Disorder which is on the rise among children. We propose a method called Auto-ASD-Network, in which we use the power of deep learning for extracting useful patterns from the data as well as discriminative power of Support Vector Machines classifier which is a very well known approach in brain disorder classification. Features extracted from the deep learning model are used as the input to the SVM classifier. In order to increase the generalizability of those features and considering the fact that deep learning methods are prone to overfitting, we employ a data augmentation method using an oversampling technique called SMOTE and double the number of items in the training set. We also use a tool called ATM in order to optimize the hyperparameters of SVM classifier using training features extracted by the deep neural network. We achieve more than 70% accuracy for 4 different datasets. Auto-ASD-Network significantly improved the results of original deep neural network (improved by 16%), SVM (improved by 26%) and state of the art classifier (improved by 14%) with the maximum accuracy of 80%.

For the future work of this study we will be focusing on designing novel deep learning based models which are able to diagnose the severity of ASD. We will also improve the performance of deep learning techniques by designing new data augmentation and simulation methods in order to increase the generalization of the deep-learning methods used in the diagnosis and the classification of mental disorders.

ACKNOWLEDGMENT

This work was supported in part by National Science Foundation grant NSF CAREER OAC 1925960. The authors would like to acknowledge hardware donation grant from NVIDIA, of a K-40 Tesla GPU which was used in the work described in this paper.

REFERENCES

[1] Alexandre Abraham, Michael P Milham, Adriana Di Martino, R Cameron Craddock, Dimitris Samaras, Bertrand Thirion, and Gael Varoquaux. 2017. Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example. *NeuroImage* 147 (2017), 736–745.

[2] Xia-an Bi, Yingchao Liu, Qin Jiang, Qing Shu, Qi Sun, and Jianhua Dai. 2018. The diagnosis of autism spectrum disorder based on the random neural network cluster. *Frontiers in human neuroscience* 12 (2018), 257.

[3] Xia-an Bi, Yang Wang, Qing Shu, Qi Sun, and Qian Xu. 2018. Classification of autism spectrum disorder using random support vector machine cluster. *Frontiers in genetics* 9 (2018), 18.

[4] Colin J Brown, Jeremy Kawahara, and Ghassan Hamarneh. 2018. Connectome priors in deep neural networks to predict autism. In *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*. IEEE, 110–113.

[5] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research* 16 (2002), 321–357.

[6] Heng Chen, Xujun Duan, Feng Liu, Fengmei Lu, Xujing Ma, Youxue Zhang, Lucina Q Uddin, and Huafu Chen. 2016. Multivariate classification of autism spectrum disorder using frequency-specific resting-state functional connectivity—a multi-center study. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 64 (2016), 1–9.

[7] John B Colby, Jeffrey D Rudie, Jesse A Brown, Pamela K Douglas, Mark S Cohen, and Zarrar Shehzad. 2012. Insights into multimodal imaging classification of ADHD. *Frontiers in systems neuroscience* 6 (2012), 59.

[8] Cameron Craddock, Yassine Benhajali, Carlton Chu, Francois Chouinard, Alan Evans, András Jakab, Budhachandra S Khundrakpam, John D Lewis, Qingyang Li, Michael Milham, et al. 2013. The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. *Neuroinformatics* (2013).

[9] R Cameron Craddock, G Andrew James, Paul E Holtzheimer III, Xiaoping P Hu, and Helen S Mayberg. 2012. A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human brain mapping* 33, 8 (2012), 1914–1928.

[10] Nicha C Dvornek, Pamela Ventola, Kevin A Pelphrey, and James S Duncan. 2017. Identifying autism from resting-state fMRI using long short-term memory networks. In *International Workshop on Machine Learning in Medical Imaging*. Springer, 362–370.

[11] Taban Eslami and Fahad Saeed. 2018. Similarity based classification of ADHD using singular value decomposition. In *Proceedings of the ACM International Conference on Computing Frontiers 2018*. ACM, 19–25.

[12] AR Jac Fredo, Afrooz Jahedi, Maya Reiter, and Ralph-Axel Müller. 2018. Diagnostic Classification of Autism using Resting-State fMRI Data and Conditional Random Forest. *Age (years)* 12, 2.76 (2018), 6–41.

[13] Xinyu Guo, Kelli C Dominick, Ali A Minai, Hailong Li, Craig A Erickson, and Long J Lu. 2017. Diagnosing autism spectrum disorder from brain resting-state functional connectivity patterns using a deep neural network with a novel feature selection method. *Frontiers in neuroscience* 11 (2017), 460.

[14] Anibal Sólon Heinsfeld, Alexandre Rosa Franco, R Cameron Craddock, Augusto Buchweitz, and Felipe Meneguzzi. 2018. Identification of autism spectrum disorder using deep learning and the ABIDE dataset. *NeuroImage: Clinical* 17 (2018), 16–23.

[15] Ehsan Hosseini-Asl, Georgy Gimel'farb, and Ayman El-Baz. 2016. Alzheimer's Disease Diagnostics by a Deeply Supervised Adaptable 3D Convolutional Network. *arXiv preprint arXiv:1607.00556* (2016).

[16] Tetsuya Iidaka. 2015. Resting state functional magnetic resonance imaging and neural network classified autism and control. *Cortex* 63 (2015), 55–67.

[17] Sarah Itani, Mandy Rossignol, Fabian Lecron, and Philippe Fortemps. 2019. Towards interpretable machine learning models for diagnosis aid: A case study on attention deficit/hyperactivity disorder. *PloS one* 14, 4 (2019), e0215720.

[18] Gajendra J Katuwal, Nathan D Cahill, Stefi A Baum, and Andrew M Michael. 2015. The predictive power of structural MRI in Autism diagnosis. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 4270–4273.

[19] Ali Khazaee, Ata Ebrahimzadeh, Abbas Babajani-Feremi, Alzheimer's Disease Neuroimaging Initiative, et al. 2017. Classification of patients with MCI and AD from healthy controls using direcfted graph measures of resting-state fMRI. *Behavioural brain research* 322 (2017), 339–350.

[20] Meenakshi Khosla, Keith Jamison, Amy Kuceyeski, and Mert Sabuncu. 2018. 3D Convolutional Neural Networks for Classification of Functional Connectomes. *arXiv preprint arXiv:1806.04209* (2018).

[21] Walter Koch, Stephan Teipel, Sophia Mueller, Jens Benninghoff, Maxmilian Wagner, Arun LW Bokde, Harald Hampel, Ute Coates, Maximilian Reiser, and Thomas Meindl. 2012. Diagnostic power of default mode network resting state fMRI in the detection of Alzheimer's disease. *Neurobiology of aging* 33, 3 (2012), 466–478.

[22] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*. 1097–1105.

[23] Hailong Li, Nehal A Parikh, and Lili He. 2018. A novel transfer learning approach to enhance deep neural network classification of brain functional connectomes. *Frontiers in neuroscience* 12 (2018), 491.

[24] Mary-Ellen Lynall, Danielle S Bassett, Robert Kerwin, Peter J McKenna, Manfred Kitzbichler, Ulrich Muller, and Ed Bullmore. 2010. Functional connectivity and brain networks in schizophrenia. *Journal of Neuroscience* 30, 28 (2010), 9477–9487.

[25] Centers of disease control and prevention. 2018. *Screening and Diagnosis of Autism Spectrum Disorder.* https://www.cdc.gov/ncbddd/autism/screening.html

[26] Sarah Parisot, Sofia Ira Ktena, Enzo Ferrante, Matthew Lee, Ricardo Guerrero, Ben Glocker, and Daniel Rueckert. 2018. Disease Prediction using Graph Convolutional Networks: Application to Autism Spectrum Disorder and Alzheimer's Disease. *Medical image analysis* (2018).

[27] Fahad Saeed. 2018. Towards quantifying psychiatric diagnosis using machine learning algorithms and big fMRI data. *Big Data Analytics* 3, 1 (2018), 7.

[28] Thomas Swearingen, Will Drevo, Bennett Cyphers, Alfredo Cuesta-Infante, Arun Ross, and Kalyan Veeramachaneni. 2017. ATM: A distributed, collaborative, scalable system for automated machine learning. In *2017 IEEE International Conference on Big Data (Big Data)*. IEEE, 151–162.

[29] Zhen Yang, Shenghua Zhong, Aaron Carass, Sarah H Ying, and Jerry L Prince. 2014. Deep learning for cerebellar ataxia classification and functional score regression. In *International Workshop on Machine Learning in Medical Imaging*. Springer, 68–76.