ASD Biomarker Detection on fMRI Images: Feature learning with Data Corruptions by Analyzing Deep Neural Network Classifier Outcomes

Xiaoxiao Li 6984086
Yale University, xiaoxiao.li@yale.edu

ASD Biomarker Detection on fMRI Images: Feature learning with Data Corruptions by Analyzing Deep Neural Network Classifier Outcomes

Xiaoxiao Li, Nicha C. Dvornek, Pamela Ventola, James Duncan,
Yale University, New Haven, CT

Background

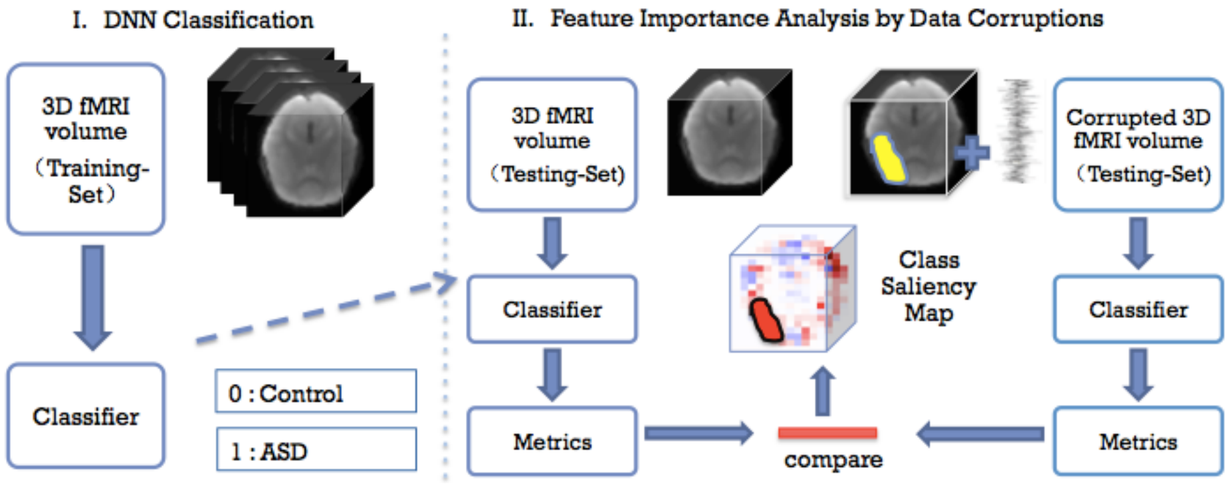
- Autism spectrum disorder (ASD) is a complex neurological and developmental disorder. Reliable biomarkers can help better target the underlying roots of ASD for diagnosis and treatment.
- Functional magnetic resonance imaging (fMRI) has helped characterize brain changes that occur in ASD [1].
- Recently deep learning was also applied to fMRI data to identify ASD [2]. However, most deep learning approaches neglected the geometric and spatial information of the whole brain 3D fMRI volume and lacked model interpretation.

Objective

- To develop a deep neural network classifier employing multi-scale kernel based representation to integrate fMRI data features for classifying ASD vs. control.
- To detect important fMRI data features activated by ASD classifier and investigate features' neurological meaning.

Methods

- **Data**: We used fMRI (146 frames) from age and IQ-matched 82 ASDs and 48 controls under *biological motion task* .
- **Data preprocessing**: We applied a sliding-window to move along the time dimension of the 4D fMRI sequence and calculated the mean and standard deviation (std) for each voxel's time series within the sliding window, thus generated mean and std 2-channel image.
- **Two-stage pipeline**: We proposed a corrupting strategy (shown in Fig. 1) to find the important regions activated by a well-trained classifier.



I. Classification : Our deep neural net (2CC3D) has 6 convolutional, 4 max-pooling and 2 fully connected layers, followed by a sigmoid output layer. The number of kernels and the layer types are denoted in each box. 0/1 stands for control/ASD. More details are shown in Fig 2

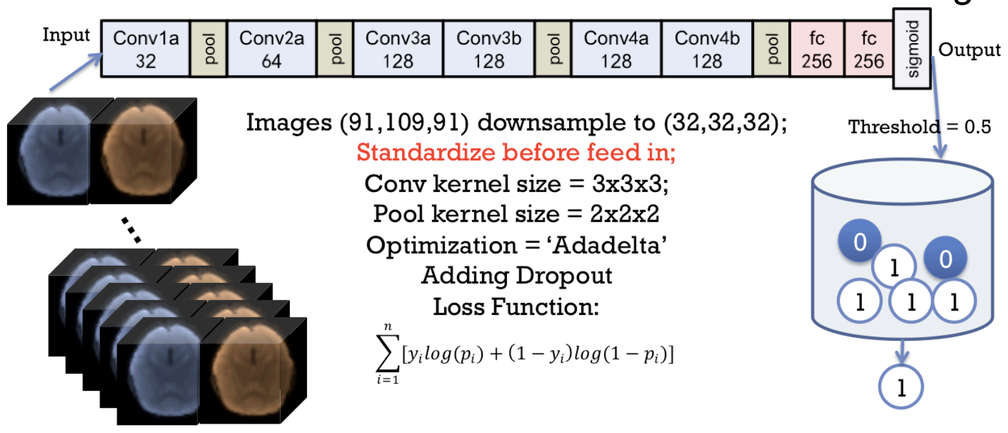


Fig 2. 2CC3D architecture and classification diagram

II. Feature Importance by Aggregating Corruptions Analysis: We parcellated the brain fMRI into 168 region of interests (ROIs). Our aim was remove data information ROI by ROI to make the corrupted fMRI. Comparing the prediction output distribution for ASDs and controls by Jensen-Shannon distance (JS), we interpret the corrupted regions which make JS smaller (difficult to classify) as important regions activated by classifier. JS distance is defined as the following equation, where P_r and P_g are two distributions.

$$\mathrm{JS}(\mathbb{P}_r,\mathbb{P}_g) = \frac{1}{2}\mathrm{KL}\left(\mathbb{P}_r \| \frac{\mathbb{P}_r+\mathbb{P}_g}{2}\right) + \frac{1}{2}\mathrm{KL}\left(\mathbb{P}_g \| \frac{\mathbb{P}_r+\mathbb{P}_g}{2}\right), \mathrm{KL}(\mathbb{P}_r\|\mathbb{P}_g) = \int \log\left(\frac{P_r(x)}{P_g(x)}\right) P_r(x)\mu(dx)$$

However, there is no certain way to fully remove the original information without adding side effects. So we aggregate the results of 1) Blackout and 2) Add Gaussian noise. The algorithm is shown:

Algorithm 1 Important Feature Detection

$\mathbb{P}^o \leftarrow f(X)$, where f is classification model
$JS_0 \leftarrow JS(\mathbb{P}^o_{ASD}, \mathbb{P}^o_{control})$
for r in ROIs **do**
  $\mathbb{P}^b \leftarrow f(X_{\setminus r})$(blackout), $\mathbb{P}^n \leftarrow f(X+\sigma_r)$(add noise)
  $JS_1 \leftarrow JS(\mathbb{P}^b_{ASD}, \mathbb{P}^b_{con})$, $JS_2 \leftarrow JS(\mathbb{P}^n_{ASD}, \mathbb{P}^n_{con})$
  $Shift^b \leftarrow \mathbb{P}^b - \mathbb{P}^o$, $Shift^n \leftarrow \mathbb{P}^n - \mathbb{P}^o$
  if $\mathbf{Cov}(Shift^b, Shift^n) > 0$, $JS_1 < JS_0$ and $JS_2 < JS_0$ **then**
    r is an important feature

Results

- **Classification**
  - We used 4 rounds of cross validation. For each round, 85%,7% and 8% subjects were selected as training data, validation data and testing data respectively.
  - We compared 2CC3D model with l2 regularized Logistic Regression, Support Vector Machine (SVM) and Random Forest (RF) . 2CC3D F−score outperformed.

($Input$ =2-$channel$, $w = 3$, $stride = 1$)

Model	Logistic	SVM	RF	2CC3D
F-score	0.69±0.14	0.68±0.06	0.82±0.06	**0.89±0.05**

Table 1. F-score of different models

- **Biomarkers**
  - Based on the shift of the new prediction distributions from corrupted data compared with the original ones, we divided the important regions into three groups by:
    - Gourp1:Shift to middle. Equally activated by the classifier for both ASD and Control;
    - Group2:Shift to control. More important for identifying ASD;
    - Group3:Shift to ASD. More important for identifying control.
  - Through Neurosynth decoding, we draw the radar chart showing the brain functional differences between ASD and control.

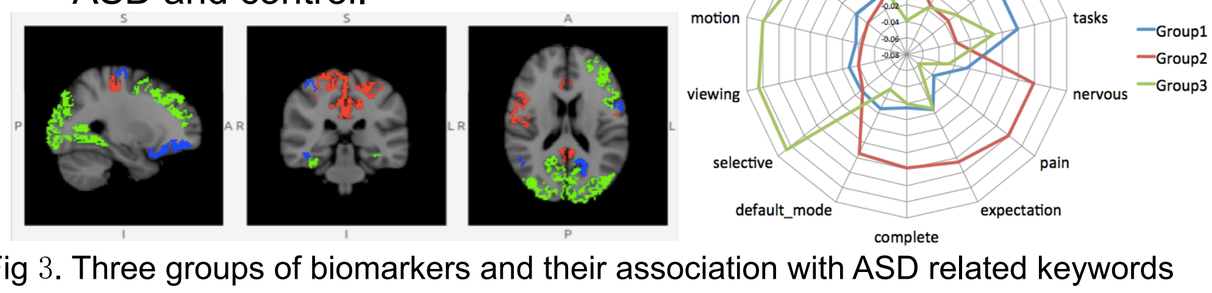


Fig 3. Three groups of biomarkers and their association with ASD related keywords

Conclusions

- Trained a deep neural network (2CC3D) classifier, which achieved high accuracy for identifying ASD and control.
- Designed an aggregating corruption method to robustly detect brain regions activated by the classifier, and found biomarkers based on the shift of prediction distribution.
- Future study will be done to
  - Improve classifier's performance;
  - Investigate how to improve the robustness of generating corrupting data and other feature importance analysis methods.

References

[1] Martha D. Kaiser, et al., "Neural signatures of autism" Proceedings of the National Academy of Sciences ,vol. 107, no. 49, pp, 2010.
[2] Yu Zhao, et al., "Automatic recognition of fmri-derived functional networks using 3d convolutional neural networks," IEEE Transactions on Biomedical Engineering , 2017