Informative Feature-Guided Siamese Network for Early Diagnosis of Autism

Kun Gao, Yue Sun, Sijie Niu, and Li Wang(✉)

Department of Radiology and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, USA
li_wang@med.unc.edu

Abstract. Autism, or autism spectrum disorder (ASD), is a complex developmental disability, and usually diagnosed with observations at around 3-4 years old based on behaviors. Studies have indicated that the early treatment, especially during early brain development in the first two years of life, can significantly improve the symptoms, therefore, it is important to identify ASD as early as possible. Most previous works employed imaging-based biomarkers for the early diagnosis of ASD. However, they only focused on extracting features from the intensity images, ignoring the more informative guidance from segmentation and parcellation maps. Moreover, since the number of autistic subjects is always much smaller than that of normal subjects, this class-imbalance issue makes the ASD diagnosis more challenging. In this work, we propose an end-to-end informative feature-guided Siamese network for the early ASD diagnosis. Specifically, besides T1w and T2w images, the discriminative features from segmentation and parcellation maps are also employed to train the model. To alleviate the class-imbalance issue, the Siamese network is utilized to effectively learn what makes the pair of inputs belong to the same class or different classes. Furthermore, the subject-specific attention module is incorporated to identify the ASD-related regions in an end-to-end fully automatic learning manner. Both ablation study and comparisons demonstrate the effectiveness of the proposed method, achieving an overall accuracy of 85.4%, sensitivity of 80.8%, and specificity of 86.7%.

Keywords: Early diagnosis · Siamese network · Subject-specific attention · Autism spectrum disorder

1 Introduction

Autism spectrum disorder (ASD), a developmental disability, would occur in all ethnic, racial, and economic groups. People suffering from ASD always perform abnormally in behaviors, communication and interaction, such as difficult communication with other people and repetitive behaviors. According to the latest report from Centers for Disease Control and Prevention, one in 54 children aged 8 years in the U.S. was diagnosed as ASD [1]. However, due to the absence of early biomarkers, ASD cannot be reliably diagnosed until around 3–4 years of age [2], only through a long process that involves observing behaviors, including language, social interactions and physical movements.

M. Liu et al. (Eds.): MLMI 2020, LNCS 12436, pp. 674–682, 2020.
https://doi.org/10.1007/978-3-030-59861-7_68

Consequently, intervention efforts may miss a critical developmental window [3]. Given the potential diagnostic instability of ASD in infancy, it is critically important to develop a system for categorizing infants at risk of ASD, who do not yet meet criteria for behaviors-based diagnosis, thus helping prevent the development of ASD [4].

Currently, there are very few works focusing on the prediction of ASD at the early stage. Shen et al. predicted ASD at 12–15 months based on the ratio of extra-axial cerebrospinal fluid (EA-CSF) to total cerebral volume [5]. They found that the ratio of 0.14 yielded 78% sensitivity and 79% specificity in predicting ASD diagnosis. In addition to EA-CSF, some works [6–8] indicated that the abnormal development of cortical gray and white matter in the infants can be used to distinguish ASD and normal control (NC). The total brain volume enlargement reported in [9–11], especially between 12 and 24 months of age, was also viewed as a crucial difference between ASD and NC. Besides, high risk-ASD infants showed a faster increasing rate in terms of cortical surface area expansion, compared with NC [10, 11]. Furthermore, Shen et al. [12] combined extra-axial CSF volume, total brain volume, age, and sex information for prediction, achieving accuracy of 78%, sensitivity of 84%, and specificity of 65%. Recently, a multi-channel convolutional neural network with a patch-level data-expanding strategy was proposed to automatically identify infants with risk of ASD at 24 months of age [13], achieving accuracy of 76.2%.

However, there are three key limitations of previous works. 1) Pre-defined brain landmarks/biomarkers isolated to the subsequent learning stage may lead to sub-optimal prediction performance due to potential heterogeneity in two standalone stages. 2) Previous works typically only relied on intensity images for classification, ignoring the more informative guidance from segmentation and parcellation maps, e.g., Li et al. [14] found many regions of interest (ROIs) with significant differences between ASD and NC groups. 3) There is a class-imbalance issue, i.e., the number of autistic subjects is far less than that of normal subjects. To address these limitations, we develop an end-to-end informative feature-guided Siamese network for the automated autism diagnosis at 24 months of age. In this work, besides T1w and T2w images, informative features from the segmentation map as well as the parcellation map are also used to train the model. Besides, inspired by [15], a subject-specific attention module is employed to locate the disease-related area in terms of ASD, which can make the prediction more reasonable. To alleviate the class-imbalance issue, we leverage the Siamese network [16] consisting of two identical subnetworks that share identical weights to learn what makes the pair of inputs belong to the same class or different classes, instead of directly classifying an input as ASD or NC in the conventional methods.

2 Dataset Acquisition and Preprocessing

There are a total of 247 subjects used in this work, including 52 ASD subjects and 195 NC subjects, which were gathered from National Database for Autism Research (NDAR) [17]. All images were acquired at 24 months of age on a Siemens 3T scanner. T1w images were acquired with parameters: TR/TE = 2400/3.16 ms and voxel resolution = $1 \times 1 \times 1$ mm^3. T2w images were acquired with parameters: TR/TE = 3200/499 ms and voxel resolution = $1 \times 1 \times 1$ mm^3. Besides T1w images and T2w images, we also

employ the segmentation map and parcellation map for the diagnosis of ASD, which were generated by a publicly available software iBEAT V2.0 Cloud (http://www.ibeat.cloud). Note that iBEAT V2.0 Cloud has been validated on 4100+ infant brain images with various protocols and scanners from 60+ institutions, including Harvard Medical School, Stanford University, Yale University, and Princeton University. For the tissue segmentation [3], each infant brain image was segmented into white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF). For the parcellation, each infant brain image was labeled into 151 ROIs, with 133 ROIs for cerebrum and 18 ROIs for cerebellum. After careful inspection, all the segmentation and parcellation maps generated by the iBEAT V2.0 Cloud passed the quality assurance. An example of T1w image, T2w image, and corresponding segmentation and parcellation maps is presented in Fig. 1. To facilitate the following diagnosis, all images/maps were rigidly aligned to an infant atlas [18] and further cropped to an identical size.

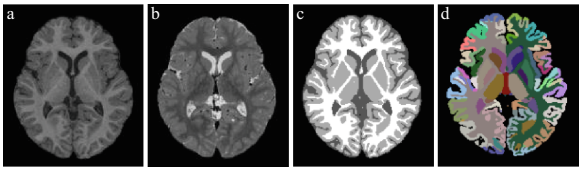

Fig. 1. An example of T1w image (a), T2w image (b), corresponding segmentation (c) and parcellation maps (d), generated by iBEAT V2.0 Cloud (http://www.ibeat.cloud).

3 Method

The proposed informative feature-guided Siamese network is shown in Fig. 2, which consists of four components: 1) feature extractor to extract discriminative features from T1w images, T2w images, segmentations and parcellation maps, 2) feature concatenation/fusion (FCF) module to fuse the extracted features, 3) subject-specific attention (SSA) module to identify ASD-related location, and 4) Siamese classifier to identify the class of the input subject.

Feature Extractor. We employ a convolution neural network as the feature extractor, to capture diverse features from T1w image, T2w image, segmentation, and parcellation maps. As a plug-in unit, the feature extractor could be implemented by any convolution neural network, such as VGGNet, ResNet or DenseNet. Besides, note that the extractors of different inputs are with the same structure but different weights. In this way, different extractors are used to learn the distinctive patterns from different inputs.

In details, there are six $3 \times 3 \times 3$ convolution layers (Conv), followed by batch normalization (BN) and parametric rectified linear unit (PReLu), and two $4 \times 4 \times 4$ downsampling layers for each extractor. The number of filters from Conv1 to Conv6 are 64, 64, 96, 128, 256, and 32, respectively. To reserve more information in the shallow stage/layer of the model, the convolution layer with stride of $4 \times 4 \times 4$ is used to downsample feature maps and increase receptive field, whereas the stride of other convolution layers is set as $1 \times 1 \times 1$.

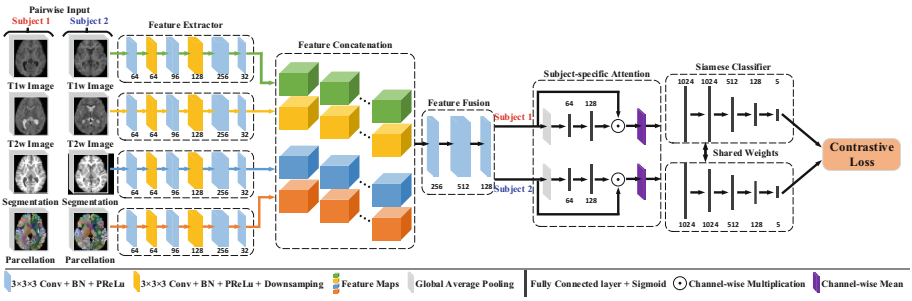

Fig. 2. Schematic diagram of the proposed end-to-end informative feature-guided Siamese network for the automated ASD diagnosis.

Feature Concatenation and Fusion Module. As shown in Fig. 2, there are four feature maps generated from the inputs, indicated by different colors. Considering that the features of each subject represent different information, we adopt concatenation operation for the informative features with the output channel, instead of element-wise addition. Subsequently, the concatenated features are fed into the fusion module for the further fusion and extraction. More specifically, the fusion module contains three convolution layers, with the kernel size of $4 \times 4 \times 4$ and the stride of $1 \times 1 \times 1$.

Subject-specific Attention Module. Previous works [9, 12, 13] have indicated the differences between ASD and NC subjects based on the MR images, such as the enlargement of total brain volume [9], the extra-axial CSF volume [12], and the detected ROIs with statistically significant difference between ASD and NC subjects. However, the predefined/detected landmarks/biomarkers are isolated to the subsequent ASD diagnosis procedure, which may lead to the sub-optimal performance. In this work, we deploy the SSA module to identify the ASD-related regions in an end-to-end fully automatic learning manner.

As shown in Fig. 2, the SSA module includes two parallel branches that consist of group average pooling, two fully connected layers, channel-wise multiplication, and channel-wise mean, where two branches with identical structure have different parameters. In details, the feature generated from FCF module is first squeezed as a vector using global average pooling to generate channel-wise statistics, regarding as a channel descriptor Z for the input feature, such that the i-th element of Z is calculated by:

$$Z_i = \frac{1}{H \times W \times L} \sum_{h=1}^{H} \sum_{w=1}^{W} \sum_{l=1}^{L} f_i^{h,w,l} \tag{1}$$

where f indicates the input features, $H \times W \times L$ is the spatial dimensions of the feature. To learn the nonlinear dependencies between channels better, a bottleneck with two fully connected layers are employed:

$$s = \delta\{W_2(\delta\{W_1 Z\})\} \tag{2}$$

where s is the scalar that will be applied to the input features, $W_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ and $W_2 \in \mathbb{R}^{C \times \frac{C}{r}}$ are the weights of the fully connected layers, r is the reduction ratio for the

bottleneck, C is the number of input features, and δ represents the activation function. Subsequently, the input feature is rescaled by s:

$$\tilde{f} = f \odot s \tag{3}$$

where $\tilde{f}$ is the rescaled feature and $\odot$ indicates the channel-wise multiplication. Finally, the attention map A is obtained by the channel-wise mean:

$$A = \frac{1}{C} \sum\nolimits_{c=1}^{C} f_c \tag{4}$$

Siamese Classifier. To alleviate the class-imbalance issue, we employ the Siamese network to identify whether the pair-wise inputs are from the same class or not. In detail, the Siamese classifier is composed of two identical subnetworks that share identical weights, in which each subnetwork consists of five fully connected layers. After each fully connected layer, the sigmoid function is used as the activation function, followed by the dropout layer to alleviate the overfitting problem. It should be noted that the sigmoid function is more suitable than PReLu function in the Siamese network as it can regularize the outputs to the same range of 0–1 to accelerate the training procedure.

In our model, the first two modules (i.e., the feature extractor and FCF module) are built to non-linearly map the input images into a low dimensional space, and then the SSA module is established to make the model focus on the ASD-related region. At last, the Siamese classifier is used to calculate the distance of the pair-wise inputs in such a way that the distance is small if the inputs belong to the same class and large otherwise.

Let X_1 and X_2 be the pair-wise attention maps from two inputs (e.g., subject 1 and subject 2 in Fig. 2). Let Y be the binary label of the pair-wise inputs, $Y = 0$ if the inputs are from the same class and $Y = 1$ otherwise. Let w be the shared weights of the two subnetworks, and $G_w()$ indicates the Siamese classifier. Then the contrastive loss $\mathcal{L}_{contrastive}$ is applied to calculate the distance of pair-wise output and formulated as:

$$\mathcal{L}_{contrastive} = (1 - Y)(D)^2 + Y\{\max(0, m - D)\}^2, D = \sqrt{\{G_w(X_1) - G_w(X_2)\}^2} \tag{5}$$

where D is defined as the distance between the outputs $G_w(X_1)$ and $G_w(X_2)$.$m > 0$ is a margin value. Having a margin indicates that dissimilar pairs that are beyond this margin will not contribute to the loss.

Hence, for the diagnosis of ASD/NC, if the class of one input is pre-defined, the class of the other input can be determined according to the contrastive loss. Compared with classifying the subject as ASD/NC directly, the proposed method, which identifies the testing subject and each training subject as the same class or not, is more effective and reasonable.

4 Experiments and Results

4.1 Implementations

In our experiments, the margin m is set as 2. The Stochastic Gradient Descent (SGD) optimizer with momentum is used during the training stage. Besides, the initial learning

rate is set as 0.001 and the cosine annealing decay is used to adjust the learning rate. In the testing phase, the distance between the given testing subject and each training subject is calculated to determine whether they are the same class or not, and the final result is obtained by majority voting. We use a 3-fold cross-validation strategy to evaluate the performance of the proposed method. More specifically, we choose two-thirds ASD subjects and two-thirds NC subjects as training set for each fold.

4.2 Ablation Study

In this part, we present the results of ablation experiments to explore the different combinations of T1w, T2w, segmentation and parcellation maps. The corresponding results in terms of sensitivity, specificity, and accuracy of each comparison and the corresponding *p*-values are listed in Table 1.

Table 1. The results of ablation comparisons. *p*-value is calculated between any input and the proposed input (T1w+T2w+Segmentation+Parcellation+Attention) with the Siamese network.

Input	Sensitivity	*p*-value	Specificity	*p*-value	Accuracy	*p*-value
T1w	76.9%	0.3389	81.5%	0.0272	80.6%	0.1810
T2w	73.1%	0.0560	80.0%	0.0418	78.6%	0.0307
T1w+T2w	75.0%	0.1377	82.1%	0.0313	80.6%	0.0311
T1w+T2w +Segmentation	82.7%	0.6481	83.1%	0.3375	83.0%	0.3515
T1w+T2w+Segmentation +Parcellation	82.7%	0.5476	84.6%	0.1347	84.2%	0.2729
Proposed method without Siamese network	71.2%	0.0417	77.9%	0.0339	76.5%	0.0192
Proposed method	80.8%	–	**86.7%**	–	**85.4%**	–

From Table 1, we can see that inputs with the segmentation map and parcellation map can improve the performance, which indicates the proposed informative guidance is effective for autism classification. In addition, we replaced the Siamese network with a conventional CNN for the finial classifier but kept others the same, with the performance listed in the last second row. It can be seen that the Siamese network can provide more accurate prediction compared with the conventional CNN. The *p*-values presented in Table 1 also show the effectiveness of the proposed method. Especially for the specificity, a statistically significant difference is reported for the result that is only trained on the intensity images (T1w, T2w, as well as T1w + T2w) and the result provided by conventional CNN (proposed method without the Siamese network), i.e., p-value < 0.05, compared with the proposed method.

Besides, we also show the attention maps generated from different inputs in Fig. 3. We can observe that the cerebellum area, which has been consistently reported in ASD [19], is highlighted in almost all attention maps. Additionally, the insular lobe is also

pointed up in Fig. 3(f2), which has a role in different aspects, such as asymbolia for pain, sensory area and motor association area, showing the consistency with characteristics of ASD, e.g., indifference to pain/temperature and restricted/repetitive behaviors. Similar findings have been also reported in [20, 21].

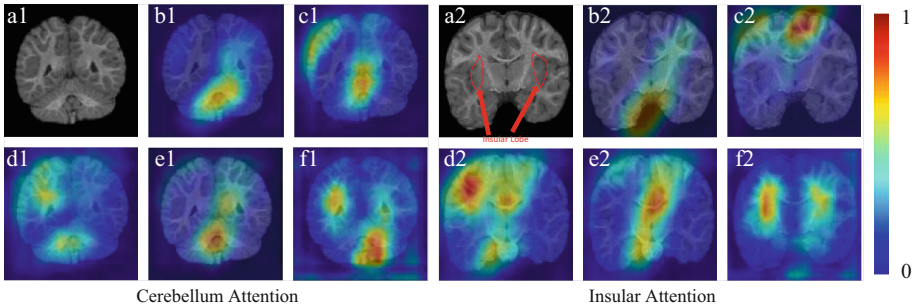

Fig. 3. Typical attention maps generated by different inputs: T1w (b1, b2), T2w (c1, c2), T1w+T2w (d1, d2), T1w+T2w+segmentation (e1, e2), and T1w+T2w+segmentation+parcellation map (f1, f2). (a1, a2) is the T1w image with different slices.

4.3 Comparison with State-of-the-Art Methods

In addition, we also compare our method with random forest (RF)-based method [22], the multi-channel CNNs [13] that utilizes the patches from anatomical landmarks to predict the status (ASD/NC) of the given subject, the EA-CSF-based method [23], and the classification method based on EA-CSF volume, total brain volume, age, and sex information (EA-CSF-BAS) [12]. Although it is hard to directly make a comparison due to different subjects involved, as shown in Table 2, our proposed method can generally achieve more than 7% increase in terms of accuracy compared with other methods [12, 13, 23], benefiting from the informative features from four input data and the Siamese classifier. Specifically, the proposed method can extract more meaningful features from different inputs, and can alleviate the class-imbalance issue due to the learning strategy of pair-wise inputs. Although our method does not perform better than the results reported in [22], our method is validated on 247 subjects, which is much more than 38 subjects involved in [22]. Furthermore, our method is fully data-driven and does not need to rely on pre-defined brain landmarks [13] or the pre-defined biomarkers, e.g., cortex shape [22], EA-CSF [23] or total brain volume [12].

Table 2. Experimental comparison with state-of-the-art methods.

	Number of ASD/NC	Modality	Age (in month)	Accuracy
RF-based method [22]	19/19	DTI	6	86.8%
Multi-channel CNNs [13]	61/215	T1w + T2w	24	76.2%
EA-CSF-based method [23]	47/296	T1w	6	69.0%
EA-CSF-BAS [12]	159/77	T1w	6	78.0%
The proposed method	52/195	T1w+T2w	24	85.4%

5 Conclusion

In this paper, we propose an end-to-end informative feature-guided Siamese network for the ASD diagnosis at 24 months of age. Specifically, we utilize the features from T1w image, T2w image, segmentation map and parcellation map to distinguish the essential difference between ASD and NC subjects. Then the SSA module is employed to locate the ASD-related region, and finally we employ the Siamese network to learn what makes the pair of inputs belong to the same class or different classes. The ablation experiments and comparison with other methods demonstrate the effectiveness of the proposed method. In the future, we will extend our work to 6-month-old infant subjects and validate on multi-site datasets.

Acknowledgements. Data used in the preparation of this work were obtained from the NIH-supported National Database for Autism Research (NDAR). NDAR is a collaborative informatics system created by the National Institutes of Health to provide a national resource to support and accelerate research in autism. This paper reflects the views of the authors and may not reflect the opinions or views of the NIH or of the Submitters submitting original data to NDAR. This work was supported in part by National Institutes of Health grants MH109773 and MH117943.

References

1. Maenner, M.J., Shaw, K.A., Baio, J.: Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2016. MMWR Surveillance Summ. **69**, 1 (2020)
2. Damiano, C.R., Mazefsky, C.A., White, S.W., Dichter, G.S.: Future directions for research in autism spectrum disorders. J. Clin. Child Adolescent Psychol. **43**, 828–843 (2014)
3. Wang, L., et al.: Volume-based analysis of 6-month-old infant brain MRI for autism biomarker identification and early diagnosis. In: Frangi, Alejandro F., Schnabel, Julia A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11072, pp. 411–419. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00931-1_47
4. Committee on Educational Interventions for Children with Autism; Board on Behavioral, C., and Sensory Sciences; Board on Children, Youth and Families; Division of Behavioral and Social Sciences and Education; National Research Council (ed.): Educating Children with Autism. National Academy Press, Washington, DC (2001)

5. Shen, M.D., et al.: Early brain enlargement and elevated extra-axial fluid in infants who develop autism spectrum disorder (2013)
6. Wolff, J.J., et al.: Neural circuitry at age 6 months associated with later repetitive behavior and sensory responsiveness in autism. Molecular Autism **8**, 8 (2017)
7. Swanson, M.R., et al.: Subcortical brain and behavior phenotypes differentiate infants with autism versus language delay. Biol. Psychiatry: Cogn. Neurosci. Neuroimaging **2**, 664–672 (2017)
8. Emerson, R.W., et al.: Functional neuroimaging of high-risk 6-month-old infants predicts a diagnosis of autism at 24 months of age. Sci. Transl. Med. **9**, eaag2882 (2017)
9. Shen, M.D., et al.: Early brain enlargement and elevated extra-axial fluid in infants who develop autism spectrum disorder. Brain **136**, 2825–2835 (2013)
10. Hazlett, H.C., et al.: Early brain development in infants at high risk for autism spectrum disorder. Nature **542**, 348–351 (2017)
11. Panizzon, M.S., et al.: Distinct genetic influences on cortical surface area and cortical thickness. Cereb. Cortex **19**, 2728–2735 (2009)
12. Shen, M.D., et al.: Extra-axial cerebrospinal fluid in high-risk and normal-risk children with autism aged 2–4 years: a case-control study. The Lancet Psychiatry **5**, 895–904 (2018)
13. Li, G., Liu, M., Sun, Q., Shen, D., Wang, L.: Early diagnosis of autism disease by multi-channel CNNs. In: Shi, Y., Suk, H.-I., Liu, M. (eds.) MLMI 2018. LNCS, vol. 11046, pp. 303–309. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00919-9_35
14. Li, G., et al.: A preliminary volumetric MRI study of amygdala and hippocampal subfields in autism during infancy. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 1052–1056. IEEE (2019)
15. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)
16. Chopra, S., Hadsell, R., LeCun, Y.: Learning a similarity metric discriminatively, with application to face verification. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), vol. 1, pp. 539–546. IEEE (2005)
17. Payakachat, N., Tilford, J.M., Ungar, W.J.: National Database for Autism Research (NDAR): big data opportunities for health services research and health technology assessment. Pharmacoeconomics **34**, 127–138 (2016)
18. Shi, F., et al.: Infant brain atlases from neonates to 1-and 2-year-olds. PLoS ONE **6**, e18746 (2011)
19. Schmitt, L.M., Cook, E.H., Sweeney, J.A., Mosconi, M.W.: Saccadic eye movement abnormalities in autism spectrum disorder indicate dysfunctions in cerebellum and brainstem. Mol. Autism **5**, 47 (2014)
20. Caria, A., de Falco, S.: Anterior insular cortex regulation in autism spectrum disorders. Front. Behav. Neurosci. **9**, 38 (2015)
21. Yamada, T., et al.: Altered functional organization within the insular cortex in adult males with high-functioning autism spectrum disorder: evidence from connectivity-based parcellation. Mol. Autism **7**, 1–15 (2016)
22. Mostapha, M., Casanova, M.F., Gimel'farb, G., El-Baz, A.: Towards non-invasive image-based early diagnosis of autism. In: Navab, N., Hornegger, J., Wells, William M., Frangi, Alejandro F. (eds.) MICCAI 2015. LNCS, vol. 9350, pp. 160–168. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24571-3_20
23. Shen, M.D., et al.: Increased extra-axial cerebrospinal fluid in high-risk infants who later develop autism. Biol. Psychiatry **82**, 186–193 (2017)