RESEARCH ARTICLE

Unified framework for early stage status prediction of autism based on infant structural magnetic resonance imaging

Kun Gao[1] | Yue Sun[1] | Sijie Niu[1,2] | Li Wang[1]

[1]Developing Brain Computing Lab, Department of Radiology and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[2]School of Information Science and Engineering, University of Jinan, Jinan, China

Correspondence
Li Wang, Developing Brain Computing Lab, Department of Radiology and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, 130 Mason Farm Road, Chapel Hill, NC 27599, USA.
Email: li_wang@med.unc.edu

Sijie Niu, School of Information Science and Engineering, University of Jinan, No. 336, West Road of Nan Xinzhuang, Jinan, China.
Email: sjniu@hotmail.com

Funding information
National Institutes of Health, Grant/Award Numbers: MH109773, MH117943

Abstract

Autism, or autism spectrum disorder (ASD), is a developmental disability that is diagnosed at about 2 years of age based on abnormal behaviors. Existing neuroimaging-based methods for the prediction of ASD typically focus on functional magnetic resonance imaging (fMRI); however, most of these fMRI-based studies include subjects older than 5 years of age. Due to challenges in the application of fMRI for infants, structural magnetic resonance imaging (sMRI) has increasingly received attention in the field for early status prediction of ASD. In this study, we propose an automated prediction framework based on infant sMRI at about 24 months of age. Specifically, by leveraging an infant-dedicated pipeline, iBEAT V2.0 Cloud, we derived segmentation and parcellation maps from infant sMRI. We employed a convolutional neural network to extract features from pairwise maps and a Siamese network to distinguish whether paired subjects were from the same or different classes. As compared to T1w imaging without segmentation and parcellation maps, our proposed approach with segmentation and parcellation maps yielded greater sensitivity, specificity, and accuracy of ASD prediction, which was validated using two datasets with different imaging protocols/scanners and was confirmed by receiver operating characteristic analysis. Furthermore, comparison with state-of-the-art methods demonstrated the superior effectiveness and robustness of the proposed method. Finally, attention maps were generated to identify subject-specific autism effects, supporting the reasonability of the predictive results. Collectively, these findings demonstrate the utility of our unified framework for the early-stage status prediction of ASD by sMRI.

Lay Summary

The status prediction of autism spectrum disorder (ASD) at an early age is highly desirable, as early intervention may significantly reduce autism symptoms. However, current methods for diagnosing young children are limited to behavioral assays. In this study, we propose an automated method for ASD status prediction at the age of 24 months that uses infant structural magnetic resonance imaging to identify neural features.

KEYWORDS
autism Spectrum disorder (ASD), deep learning algorithm, early-stage status prediction, infant structural MRI, subject-specific autism attention

INTRODUCTION

Autism spectrum disorder (ASD) is commonly described as a complex neurodevelopmental disorder that seriously damages the sociality and communication ability of patients. In particular, individuals diagnosed with ASD struggle with communication and language. A recent report from Centers for Disease Control and Prevention

(Maenner et al., 2020) indicates that 1 in 54 children aged 8 years has been diagnosed with ASD in the United States, with 4.3 times higher ASD prevalence among boys than girls (29.7% vs. 6.9%). Diagnosis of ASD at an early age is highly desirable, as early intervention may significantly reduce autism symptoms (Dillenburger, 2014). However, early diagnosis is challenging because of the lack of biomarkers to detect children either with or at-risk of autism during the first postnatal years of life. Diagnosis currently relies on behavioral observations that arise after birth; consequently, autism is not typically diagnosed until about 2 years of age in the U.S. (Pierce et al., 2019), when the window of opportunity for effective intervention may have already passed (Committee on Educational Interventions for Children with Autism; Board on Behavioral, 2001). Thus, in order to accelerate ASD diagnosis, it is of great importance to identify imaging-based biomarkers for early diagnosis and intervention that are independent of behavioral diagnostic symptoms of autism.

Recently, neuroimaging technology has been widely employed in the field of brain disease analysis for conditions such as ASD (Fortin et al., 2018; Li Wang et al., 2018) and Alzheimer's disease (Lian et al., 2019; Liu et al., 2020). There are several commonly used modalities for the study of brain diseases, including structural magnetic resonance imaging (sMRI), functional MRI (fMRI), positron emission tomography (PET), and diffusion tensor imaging (DTI). Researchers have leveraged characteristics of these modalities to explore changes in brain function or structure related to brain diseases. For instance, Saeed et al. (2019) proposed ASD-DiagNet to classify subjects as ASD or normal control (NC) using fMRI from subjects over 10 years of age. Furthermore, Chen et al. utilized whole brain functional connectivity networks from fMRI data as features to classify subjects (12–18 years old) as ASD or NC (Chen et al., 2016). By exploring the amygdala-centered functional connectivity during explicit and implicit threat processing, Chen et al. (2021) trained a linear kernel support vector machine (SVM) classifier to predict diagnostic value of ASD. However, most of these fMRI-based studies have focused on subjects older than 5 years of age. The application of fMRI for infants poses challenges: (1) the imaging time of fMRI is relatively long, for example, 40–55 min (Yale-Medicine, 2021); (2) for infants, it is difficult to acquire fMRI data during a resting stage; and (3) the existing fMRI protocols are mainly based on adult cohorts and cannot be easily adapted and applied to infants (Zhang et al., 2019). By contrast, sMRI has the following advantages: (1) it is faster, for example, 20–30 min (Yale-Medicine, 2021) and thus more reliable than fMRI; and (2) there are many sMRI imaging protocols specifically designed for infants, for example, BCP (Howell et al., 2019) and dHCP (Makropoulos et al., 2018). Therefore, sMRI-based approaches have

increasingly received attention in the field of early status ASD prediction. For example, based on the ratio of extra-axial cerebrospinal fluid (CSF) volume to total cerebral volume, Shen et al (2013) predefined a threshold to distinguish ASD subjects from the NC group at the age of 12–15 months. Furthermore, by combining extra-axial CSF volume, brain volume, age, and sex, Shen et al. (2018) constructed a RUSBoost classifier to predict subjects as ASD or NC at 2–4 years of age based on sMRI. In addition, Hazlett et al. (2017) investigated abnormal early brain changes, such as hyper expansion of the cortical surface area and overgrowth of the total brain volume to distinguish ASD using infant sMRI. Leveraging the ASD-related landmarks detected by a landmark discovery algorithm (Zhang et al., 2016), Li et al. (2018) proposed a multi-channel convolutional neural network (CNN) based on sMRI at the age of 24 months for ASD prediction. Additionally, Xiao et al (2017) employed a random forest classifier and utilized the regional average cortical thickness feature for ASD prediction. Similarly, Conti et al. (2020) constructed a linear-kernel SVM to explore possible ASD-specific brain structural features, concluding that a set of cortical thickness features can achieve the best predictive performance. Previous studies indicate that autistic subjects show volume differences compared with typically developing subjects, such as smaller cerebellum volume (Rogers et al., 2013; Stoodley, 2014), larger amygdala volume (Nordahl et al., 2020), or larger hippocampus volume (Xu et al., 2020). These studies are unique in identification of early biomarkers from sMRI, encouraging further study to explore the relationship between ASD and sMRI modality. However, prior investigation based on sMRI has the following key limitations: (1) previous studies have always selected/extracted features from intensity MRIs, ignoring potential meaningful features for ASD prediction that could be identified from segmentation and parcellation maps, such as the volume of specific brain structures (Nordahl et al., 2020; Rogers et al., 2013); (2) most of the previous work has relied on predefined biomarkers/landmarks, which are isolated to subsequent learning-based approaches and may lead to suboptimal predictive performance; and (3) the generalization ability of previous work may suffer from data variability related to different imaging protocols/scanners (Fortin et al., 2017; Fortin et al., 2018). We hypothesize that improved approaches could identify more reliable imaging-based biomarkers from sMRI and potentially increase the robustness of sMRI-based ASD prediction.

In this study, to alleviate the limitations and improve early diagnosis of ASD, we developed an end-to-end data-driven automated method for ASD status prediction at the age of 24 months based on sMRI. Instead of relying on intensity images directly (e.g., T1w images), we employed segmentation and parcellation mapping to alleviate inter-site data heterogeneity. In addition, given the different developmental patterns between males and

females (Postorino et al., 2015), we integrated sex information into our model. Specifically, we used a CNN to extract features from segmentation and parcellation maps, followed by a subject-specific autism attention module to identify regions related to ASD. Finally, we employed a Siamese network (Chopra et al., 2005) to calculate the contrastive loss based on distances between pairwise inputs (as defined by the data class), selected to circumvent class-imbalance issues. Herein, we present an overview of our method, including a dataset with preprocessing steps, experimental implementation, and details of the training and testing stages, as well as the results of ablation studies and comparisons with state-of-the-art methods. Notably, the results of cross-site experimentation confirm the stable generalization ability of the proposed approach. In addition to quantitative analyses to evaluate the discriminative ability of our approach, we present an attention map generated by our method to improve the reasonability of the prediction. These results verify the feasibility of this method for use as a quantitative early predictive tool for ASD that has potential utility independent of behavioral observation.
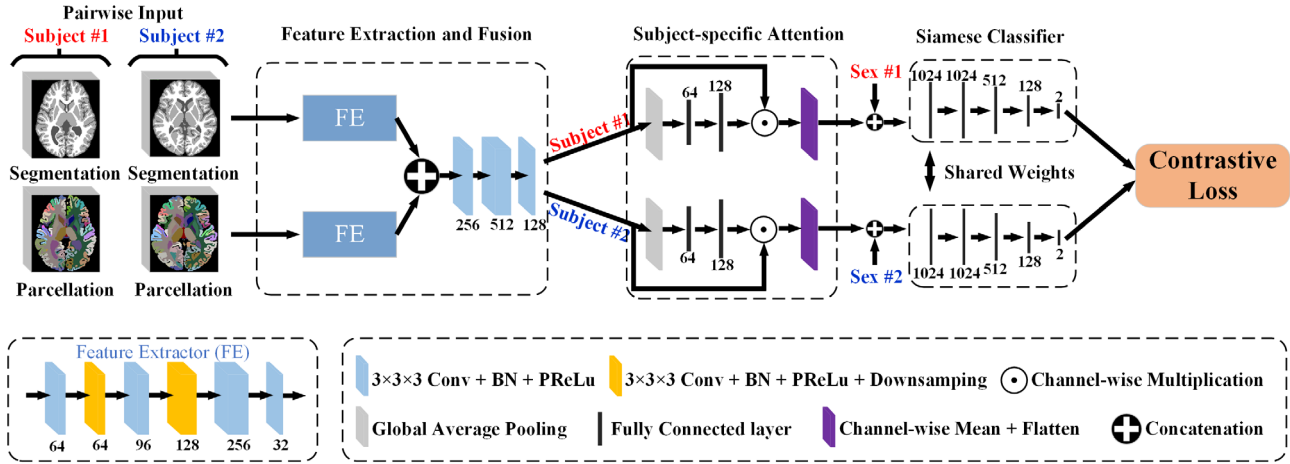
METHODS

Overview of the study design

Figure 1 illustrates the overall framework for our study. Intensity inhomogeneity correction, skull stripping, segmentation and parcellation were applied sequentially to original raw intensity images, leveraging the publicly available software iBEAT V2.0 Cloud (http://www.ibeat.cloud/). The iBEAT V2.0 Cloud has successfully processed more than 12,400+ infant brain images with different protocols and scanners from 90+ institutions. The segmentation and parcellation maps were used as input for the deep learning-based predictive model (including informative feature- and attention-guided Siamese networks). Specifically, a Siamese network was used as the main architecture to achieve early-status prediction of ASD. It should be noted that the Siamese network is one of the deep neural networks, but it is different from traditional CNNs: the Siamese network is used to find the similarity or dissimilarity of pairwise subjects by comparing their feature vectors, whereas traditional CNNs directly classify a single subject as a specific category. Additional definition of terms is provided in Table S1. Our framework included parallel paths for the training and testing phases. In the training phase, pairwise input that was randomly selected from the training set was fed into the network to optimize the weights of the model; in the testing phase, the mean distance between a given testing subject and all training subjects was calculated to distinguish whether the testing subject was from the ASD or NC group.

The details of our informative feature- and attention-guided Siamese network are shown in Figure 2. There are three components in this network: the feature extraction (FE) and fusion module, the subject-specific autism attention module, and the Siamese distinguishing module. Conventional methods directly utilize intensity images or features/biomarkers extracted from intensity images for the ASD screening; these parameters are affected by high level inter-site data heterogeneity caused by different imaging protocols/scanners. Therefore, instead of directly relying on image intensity, segmentation and parcellation maps were employed for the prediction of ASD. The infant-dedicated pipeline, iBEAT V2.0 Cloud, employs site-independent prior knowledge, including an anatomy-guided CNN for segmentation/parcellation. After segmentation and parcellation, a CNN was employed to extract features from the segmentation and parcellation maps. The extracted features were further concatenated and fused. Subsequently, a subject-specific autism attention module was used to identify
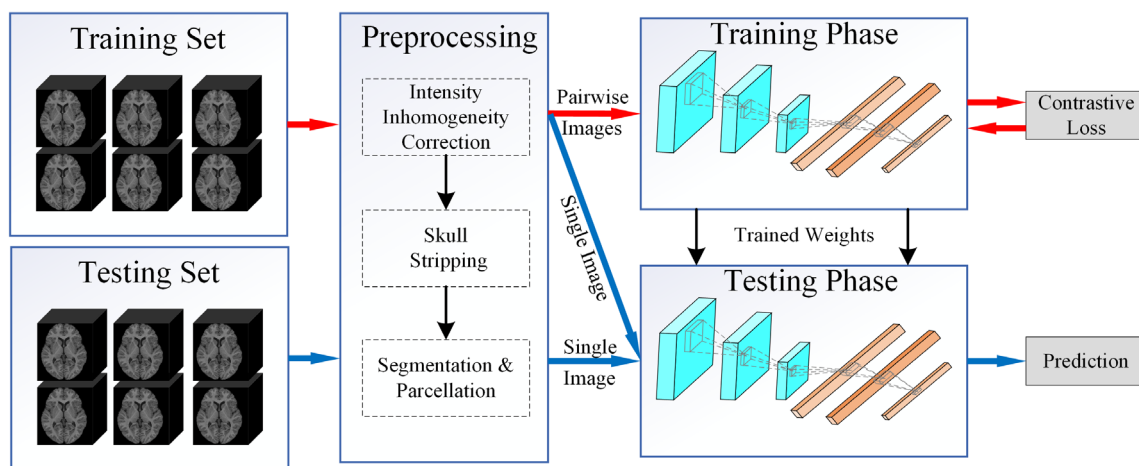
FIGURE 1 Overview of the study framework. Parallel paths were used to process raw data from a training set and a testing set from iBEAT V2.0 Cloud. After optimization with the training set, the testing set was used to distinguish autism spectrum disorder and normal control groups

FIGURE 2 The pipeline of the deep learning method for automatic autism spectrum disorder (ASD) diagnosis. (a) Segmentation and parcellation maps were obtained from iBEAT V2.0 and applied as parallel attention paths. (b) Convolutional neural network was used for feature extraction and fusion; (c) Subject-specific autism attention module was used to identify ASD-associated features; (d) Siamese classifier was used to integrate sex information and develop a matrix for calculating the contrastive loss

regions related to ASD, followed by integration of sex information. Finally, a Siamese network was deployed to determine whether pairwise subjects were from the same or different classes. More details about each of the three modules are provided below.

Feature extraction and fusion module

Two CNNs were employed to separately extract features from the segmentation and parcellation maps. Notably, the feature extractors for the segmentation and parcellation maps have identical structures but different weights. In this way, the network is encouraged to learn different features from the two maps. Next, the extracted features were combined and further fused through three convolutional layers. All convolutional layers used $3 \times 3 \times 3$ kernels with zero padding, followed by batch normalization and rectified linear unit activation. The strider of the extraction layer (marked as blue rectangles) was set as 1, while that of down sampling (marked as yellow rectangles) was set as 2 to enlarge the receptive field and alleviate information loss. To minimize the number of channels, the channel numbers of the last layer in the FE and fusion module were set as 32 and 128, respectively.

Subject-specific autism attention module

The first FE and fusion module provides features extracted from the segmentation and parcellation maps for subsequent tasks. However, some features may be redundant for the prediction task. Thus, based on the method of Hu et al. (2020), a subject-specific autism attention module was employed to identify regions

associated with ASD in an end-to-end manner. More specifically, the pairwise inputs (i.e., subject #1 and subject #2) were separately fed into two parallel attention paths. Each attention path consisted of a squeeze and excitation (SE) block (Hu et al., 2020) and a channel-wise mean, where the SE block was used to weight all features and the channel-wise mean generated the attention map for each subject. After the subject-specific autism attention module, sex information was concatenated with the features. Similar to the FE block, the two parallel paths of the subject-specific autism attention module had the same structure but different weights.

Siamese distinguishing module

Currently, most CAD systems for ASD prediction rely on traditional classifiers, such as random forest (Katuwal et al., 2018; Mostapha et al., 2015) and CNN (Fortin et al., 2018; Saeed et al., 2019) classifiers. These models usually directly classify the subjects as ASD or NC, rather than considering the distance between intra-class and inter-class variation. Instead, the Siamese network was employed to construct a relationship matrix and calculate the distance between ASD and NC using contrastive loss; thus, distances among the same class should be small while distances between different classes should be large. The contrastive loss was formulated as shown in Equation 1,

$$\mathcal{L} = (1-Y)(D)^2 +$$
$$Y\{\max(0, m-D)\}^2, D = |G_w(x_1) - G_w(x_2)| \quad (1)$$

where x_1 and x_2 indicate the features from subject #1 and subject #2 respectively, G_w represents the Siamese

network, Y is the binary label of the pairwise inputs ($Y = 0$ if the inputs are from the same class and $Y = 1$ otherwise), D is the distance between the outputs $G_w(x_1)$ and $G_w(x_2)$, and m is the margin value, which is non-negative and indicates the maximum contribution to the final loss.

Similar to the subject-specific autism attention module, the Siamese distinguishing module was composed of two parallel paths corresponding to pairwise inputs (i.e., subject #1 and subject #2, Figure 2). Each path of the Siamese network included five fully connected layers, followed by batch normalization and rectified linear unit activation. There are two advantages of the Siamese network compared with traditional CNNs. First, training samples are enlarged for the Siamese network, as any unique pairwise samples in the training set are meaningful for the Siamese network. Second, the testing phase of the Siamese network can be viewed as an ensemble system, which may improve the accuracy of the prediction.

Our informative feature- and attention-guided ASD prediction model was implemented in a single NVIDIA GTX TITAN (12GB) GPU by the PyTorch framework. In the training stage, the minibatch size was set as 2 (pairwise input) and 0.4 dropout was applied to each fully connected layer in the Siamese network. Contrastive loss and stochastic gradient descent with momentum were used to optimize the weights of the model, where the learning rate was initialized as 1e−3 with the cosine annealing decay. In the testing stage, the distance between the testing subject and each training subject was calculated to distinguish whether they were from the same class or different classes, and the mean distance between the testing subject and all training subjects was used as the result.

RESULTS

Models trained on one dataset with a set of specific imaging parameters perform poorly at other datasets with different imaging parameters

To evaluate the potential of infant sMRI to be used as a predictive method for ASD, we used two datasets that were acquired by different scanners and imaging protocols, as detailed in Table 1. They are available in the National Database for Autism Research (NDAR)

(https://nda.nih.gov). Specifically, dataset A consists of 247 subjects from the Infant Brain Imaging Study (IBIS) network (https://ibis-network.com) (Hazlett et al., 2017), and dataset B consists of 35 subjects from the Autism Centers of Excellence (ACE) (Hall et al., 2012). All images were scanned from infants at about 24 months of age without significant difference between the ages of the ASD and NC groups (*p*-value >0.05). For both datasets A and B, the infants were naturally sleeping with ear and head protection when the images were acquired. Table 1 shows that there were large differences between the two datasets in terms of scanners and imaging parameters, resulting in different image contrasts/appearances/patterns, as shown in Figure 3. These results emphasize that models trained on one dataset with a set of specific imaging parameters tend to perform poorly for other datasets with different imaging parameters (protocols/scanners), which is consistent with previous observations (Sun et al., 2021).

Segmentation and parcellation maps improve the robustness of the model for structural magnetic resonance imaging neurological image interpretation

Given the poor cross-institutional resolution of neural images trained at a single site, we sought to develop a unified framework that can be used across different sites. All intensity images were preprocessed using intensity inhomogeneity correction, skull stripping, segmentation and parcellation, leveraging the publicly available software iBEAT V2.0 Cloud (http://www.ibeat.cloud). Each infant brain image was segmented into three tissues (i.e., white matter, gray matter, and cerebrospinal fluid) and parcellated into 151 ROIs (133 ROIs in cerebrum and 18 ROIs in cerebellum). All segmentation and parcellation maps were linearly aligned to the infant atlas (Shi et al., 2011) by the FLIRT method (Jenkinson et al., 2002) and further cropped to an identical size to facilitate the subsequent learning stage.

We first used a 10-fold cross-validation strategy for dataset A to evaluate the performance of our model on a single dataset. For each fold, nine-tenths of ASD subjects and nine-tenths of NC subjects were used as a training set, while the remaining cases were used as the testing set. We compare the performance in terms of different

TABLE 1   Detailed information on the datasets used in our work

	Category	Number	Sex (M/F)	Age (months)	Scanner	TR/TE (ms)	Resolution (mm^3)
Dataset A	ASD	52	41/11	24.1 ± 0.73	Siemens (3T)	2400/3.16	1.0 × 1.0 × 1.0
	NC	195	109/86	24.3 ± 0.89			
Dataset B	ASD	22	18/4	25.1 ± 1.59	GE (3T)	6.496/2.796	0.9375 × 0.9375 × 0.9375
	NC	13	4/9	24.6 ± 1.94			

FIGURE 3 Large inter-site data heterogeneity for two infant subjects acquired by different scanners. The left subject from the Infant Brain Imaging Study network was acquired by a Siemens 3T scanner, whereas the right subject from Autism Centers of Excellence was acquired by a GE 3T scanner. The large inter-site data heterogeneity limits the application of models trained on a dataset with a set of specific imaging parameters to other datasets with different imaging parameters (protocols/scanners)
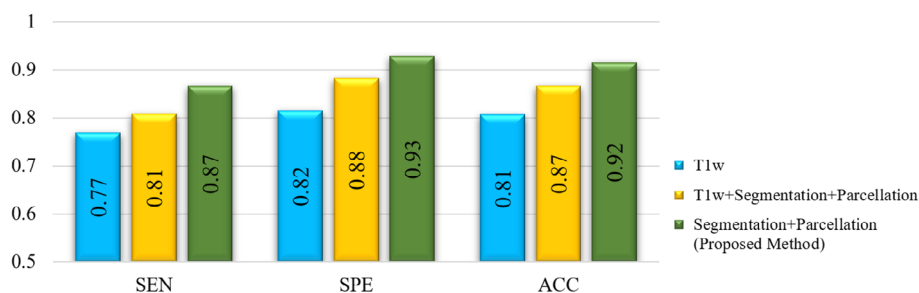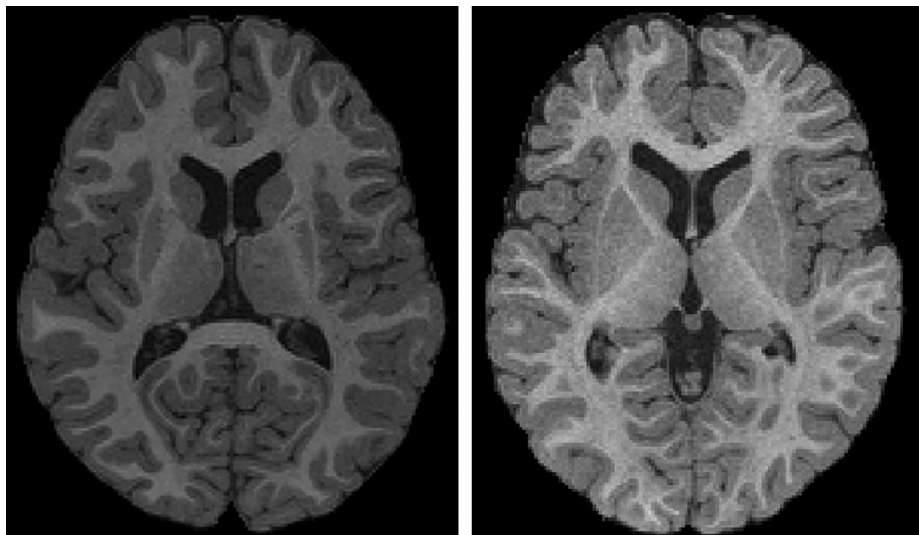
FIGURE 4 Assessment of different combinations of T1w images, segmentation, and parcellation maps. The methods, each of which was trained on dataset A, were evaluated in terms of sensitivity (SEN), specificity (SPE), and accuracy (ACC). The metric shows the performance of the model trained with T1w intensity images (the first bars), a combination of intensity images, segmentation and parcellation maps (the second bars), and the proposed segmentation + parcellation maps (the third bars)
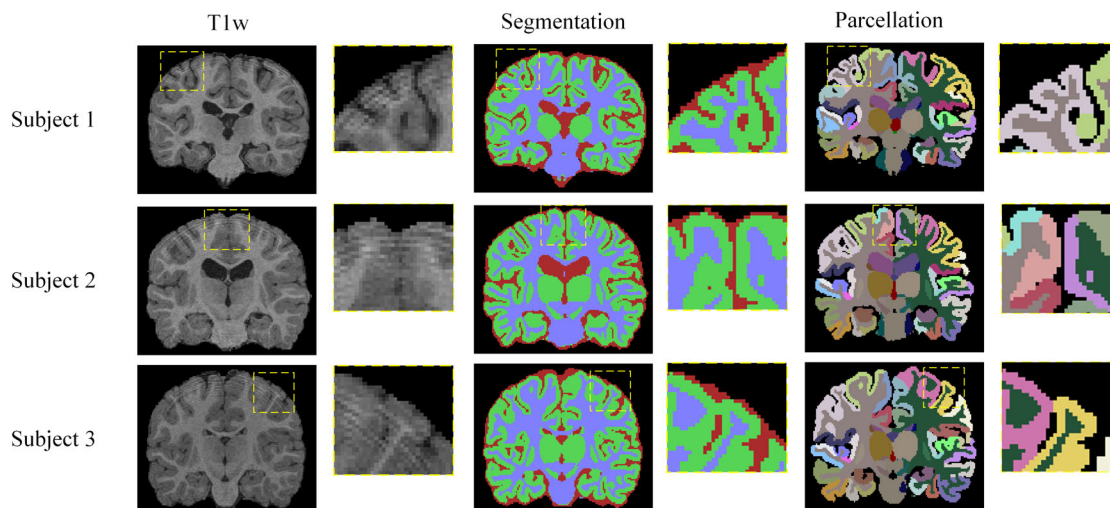
combinations of T1w images, segmentation maps, and parcellation maps on dataset A, as shown in Figure 4. The predictive performance was quantitatively evaluated by three criteria: sensitivity (SEN), specificity (SPE), and accuracy (ACC). The results demonstrate that training the model with only T1w images yielded 76.9% SEN, 81.5% SPE, and 80.6% ACC. With a combination of T1w images, segmentation and parcellation maps (the second bar for each metric), there was about 4% improvement in terms of SEN and more than 6% improvement in terms of SPE and ACC, indicating that the segmentation and parcellation maps are useful for distinguishing ASD from NC subjects. Furthermore, when the model was trained only with the proposed segmentation + parcellation maps (the last bar for each metric), more than 9% improvement was achieved in terms of SEN, SPE, and ACC. These results suggest that our method of using segmentation + parcellation maps improves the predictive performance of sMRI neurological images.

The utility of intensity images is limited because they are often corrupted by imaging noise and artifacts (Wang et al., 2019). Moreover, if test subjects are acquired with different imaging protocols/scanners, the large inter-site data heterogeneity presented in the intensity images

further degrades the performance. Therefore, to further evaluate the effect of segmentation and parcellation in improving the clarity of sMRI images from different sources, we examined representative individual images acquired at different sites. As shown in Figure 5 (the first column), the varying noise and artifacts (e.g., motion and Gibbs) across different subjects (from dataset A) confuse the classifiers and affect the predictive performance. Compared with the intensity images, the segmentation and parcellation maps generated by iBEAT V2.0 Cloud with the guidance of prior anatomy knowledge minimize these artifacts (the second and third columns). Therefore, these results suggest that the proposed segmentation + parcellation method minimizes potential artifacts caused by inter-site variability.

Integration of sex information improves the predictive ability of the proposed infant structural magnetic resonance imaging interpretation method

After the first case of autism described in the 1940s, a skewed sex ratio was demonstrated, indicating that

FIGURE 5 T1w images, segmentation maps, and parcellations maps for three representative subjects from dataset A. The first column is the T1w images affected by imaging noise and Gibbs artifacts. The second and third columns are the segmentation map and parcellation map, respectively. The images in the last two columns were generated by iBEAT V2.0 cloud with the guidance of prior anatomy knowledge

autism is significantly more common in boys than girls (Anello et al., 2009; Lai & Szatmari, 2020; Mottron et al., 2015). Recent research (Fortin et al., 2018; Shen et al., 2018) further indicates that sex information should be factored into the analysis of autism. Therefore, we explored how sex information affects our predictive model. Sex information for the images was combined with the features extracted from segmentation and parcellation maps and was used as the input in a Siamese network for the final decision. Table 2 shows the results of experiments on dataset A with 10-fold cross-validation. The model trained without the sex information achieved 84.6%, 90.3%, and 89.1% in terms of SEN, SPE, and ACC; integration of the sex information improved the predictive performance of our model, with increases of 1.9% (to 86.5%) for SEN, 2.5% (to 92.8%) for SPE, and 2.4% (to 91.5%) for ACC. To further evaluate the robustness of our method, we applied the model trained on dataset A to dataset B. As shown in Table 3 on dataset B, when trained with the sex information, the proposed model showed a similar improvement in generalization compared with the model trained without the sex information. These findings indicate that sex information improves the predictability of our approach.

To further examine the predictability of this method, we performed receiver operating characteristic curve (ROC) and area under the ROC curve (AUC) analyses (Figure 6). For dataset A, the model trained without the sex information (green) had an AUC of 87%; however, the model trained with the sex information (orange) revealed improvement in both the predictive performance and the discrimination capacity for distinguishing between ASD and NC (AUC 91%). Similarly, for dataset B, our model trained with the sex information performed better than the model without the sex

TABLE 2 Comparison of the sensitivity (SEN), specificity (SPE) and accuracy (ACC) for the analysis of dataset A without/with the sex information

|	SEN	SPE	ACC
Proposed method without sex	0.846	0.903	0.891
Proposed method with sex	0.865	0.928	0.915

TABLE 3 Comparison of the sensitivity (SEN), specificity (SPE) and accuracy (ACC) for the analysis of dataset B without/with the sex information

	SEN	SPE	ACC
Proposed method without sex	0.727	0.923	0.800
Proposed method with sex	0.818	0.846	0.829

information, though the difference was comparatively less for this dataset (AUC 86% versus 85%). These results verify that sex information is useful for qualifying the screening of ASD and NC subjects.

The proposed approach for infant structural magnetic resonance imaging outperforms other state-of-the-art methods

To further evaluate the utility of the proposed method, we directly compared it with two state-of-the-art methods: the EA-CSF method (Shen et al., 2013) and the EA-CSF-BAS method (Shen et al., 2018). The former method utilizes predefined landmarks/biomarkers, that is, the ratio of extra-axial cerebrospinal volume to whole brain volume, to distinguish ASD and NC groups; while
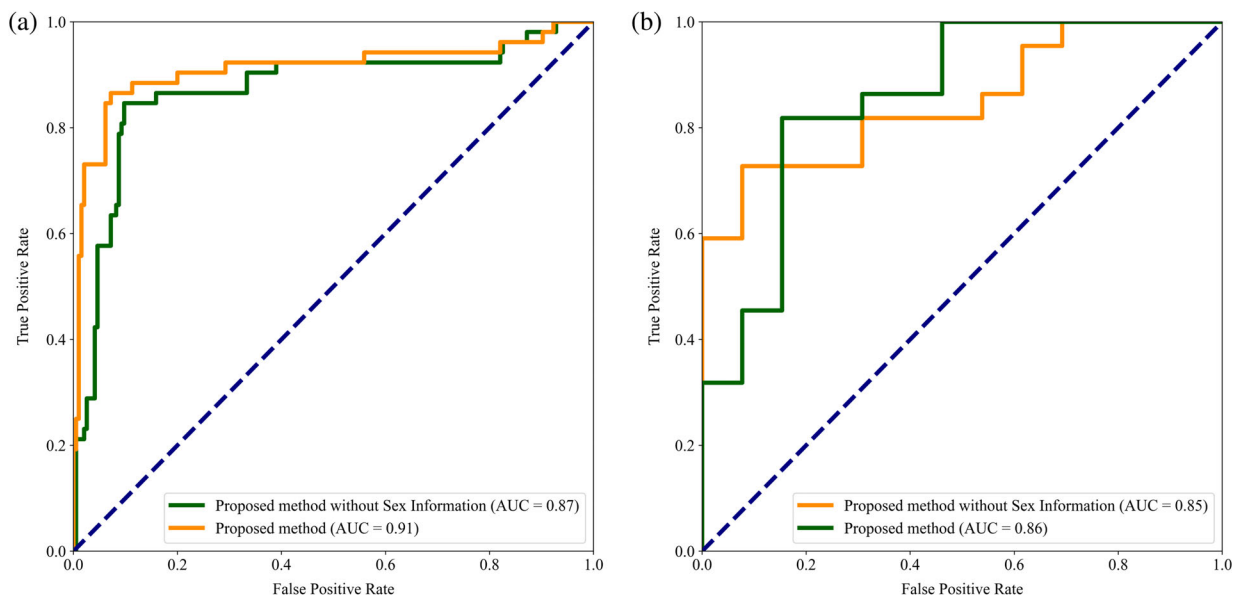
FIGURE 6 Receiver operating characteristic curves for datasets A and B with/without the sex information. Green lines represent models trained without sex information, and orange lines represent models trained with sex information. (a) Was generated by 10-fold experimental cross-validation of dataset A, while (b) was generated by applying the model trained on dataset A to dataset B directly

TABLE 4 Comparison of the proposed method with state-of-the-art methods for datasets A and B. "()" represents the p-value in comparison with the proposed method

|	Dataset A			Dataset B		
	SEN	**SPE**	**ACC**	**SEN**	**SPE**	**ACC**
EA-CSF (Shen et al., 2013)	0.731 (0.0017)	0.759 (0.0012)	0.753 (0.0023)	0.636 (0.0042)	0.538 (0.0001)	0.600 (0.0014)
EA-CSF-BAS (Shen et al., 2018)	0.788 (0.013)	0.815 (0.0059)	0.810 (0.03)	0.773 (0.37)	0.538 (0.0001)	0.686 (0.0027)
Proposed method	0.865	0.928	0.915	0.818	0.846	0.829

the latter method considers additional factors, such as age and sex. We implemented these two methods on dataset A by the same 10-fold cross-validation as for our method. For the EA-CSF method, we calculated the ratio of extra-axial fluid to total cerebral volume (fluid: brain) for each subject to determine which threshold best predicted ASD. Similar to the results of Shen et al. (2013), we determined that the ratio of fluid to brain (0.14) obtained a peak sensitivity of 73.1% and a peak specificity of 75.9% (Table 4, the first row). For the EA-CSF-BAS method, we implement the RUSBoost classifier using the RUSBoostClassifier package in PyTorch. For optimal performance, the number of estimators was set at 500, the learning rate was set at 1.0, and the algorithm was set at SAMMER. The performance in terms of SEN, SPE, and ACC was slightly higher than that of the EA-CSF method (Table 4, the second row). However, our proposed method performed better than either of the other methods, achieving more than 10% improvement in terms of the SPE and ACC for dataset A (Table 4, the third row). A similar improvement in performance was observed for dataset B,

although this dataset showed poorer performance across the board (Table 4). Thus, these findings verify the superior robustness of our method relative to other state-of-the-art methods.

Attention mapping of infant structural magnetic resonance imaging images supports known mechanisms of autism spectrum disorder

To further evaluate our method and explore the ROIs related to ASD, we generated attention maps using a subject-specific autism attention module (Figure 7). The amygdala, hippocampus, and cerebellum were highlighted by our method, which is consistent with the results of previous studies (Nordahl et al., 2020; Rogers et al., 2013; Stoodley, 2014). Specifically, abnormalities of cerebellar function are known to result in specific symptoms of ASD, such as deficits in cognitive and motor behavior (Rogers et al., 2013; Stoodley, 2014). Moreover, the amygdala and hippocampus have been shown to differ between people with and without ASD,
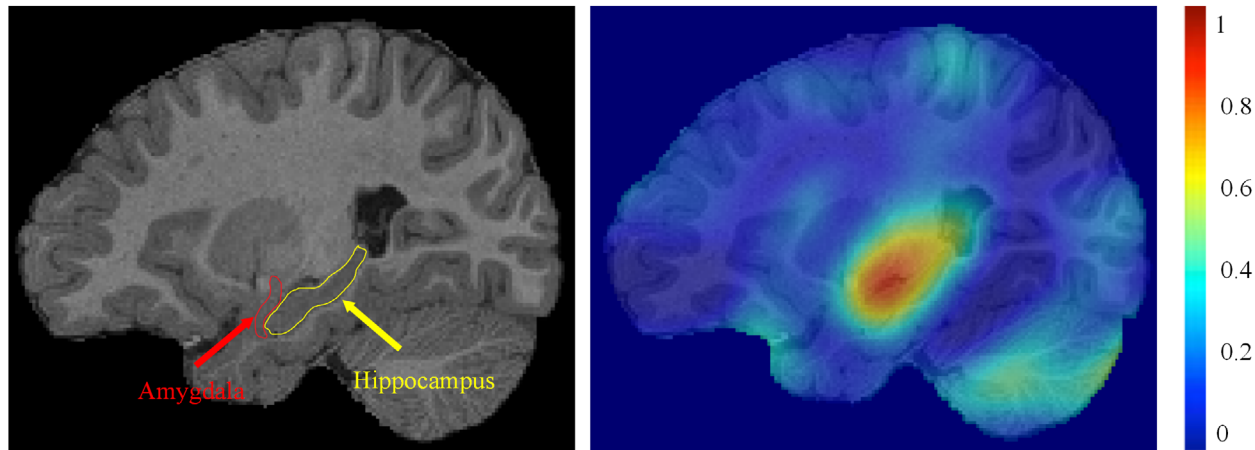
FIGURE 7 Attention maps generated by the proposed method. The amygdala, hippocampus, and cerebellum were identified by our method as abnormal regions associated with autism spectrum disorder, shown in the first and second rows respectively. In the right panels, red and blue indicate high and low discriminative power, respectively

with a smaller amygdala (Nordahl et al., 2020) or enlarged hippocampus (Xu et al., 2020) being characteristic of ASD. These results further support our method for infant sMRI interpretation, thus verifying the reliability of the proposed method.

DISCUSSION

In this study, we evaluated a novel method for infant sMRI interpretation based on segmentation and parcellation maps using two datasets. For the ablation study, we first explored different combinations of T1w, segmentation mapping, and parcellation mapping. The results indicate that models trained with segmentation and parcellation maps perform better than the model trained with intensity images. Given the different developmental patterns between males and females, we further evaluated the effect of integrating sex information into our study. The model showed better performance when trained with the sex information, indicating that sex information is a vital factor in the study of autism. We generated ROC curves to further evaluate the performance, which verified that the model was more stable when trained with the sex information, with improvements of 4% and 1% in terms of the AUC for dataset A and dataset B, respectively.

As further verification, we compared our method with other state-of-the-art methods. The results indicate the effectiveness of our method, which accurately predicted more subjects for both the ASD and NC groups, with higher SPE/SEN/ACC values. It should be noted that these competing methods are based on predefined landmarks/biomarkers. Such predefined markers may lead to suboptimal predictive performance due to the isolation between FE and the final decision. In contrast, our method is performed in a fully automated and end-to-end

manner. Furthermore, the pairwise training strategy of the Siamese network can better establish the relationship between ASD and NC with a relatively small number of subjects. Therefore, based on these advantages, our method has greater potential to achieve better performance compared with prior approaches. Results of the trained models on dataset B support these findings and further demonstrate the robustness of the proposed method, indicating that the proposed framework may have broader applicability for the clinical diagnosis of ASD.

In addition to quantitative analysis, attention maps are useful in identifying regions related to ASD. We generated an attention map using a subject-specific module and identified three regions associated with ASD as highlighted by the proposed method, including the amygdala, hippocampus, and cerebellum, which is consistent with the results of previous studies (Nordahl et al., 2020; Rogers et al., 2013; Stoodley, 2014; Xu et al., 2020). Compared with the EA-CSF method (Shen et al., 2013) and EA-CSF-BAS method (Shen et al., 2018), each of which employs predefined biomarkers, the attention map captured by our proposed method provides data-driven ROI identification, indicating the reasonability of the predictive result.

Despite the promising results presented in this study, our method has a number of potential limitations. Chiefly, our approach depends upon a public infant processing pipeline iBEAT V2.0 Cloud. In our study, the segmentation and parcellation performed by the iBEAT V2.0 Cloud improve the robustness of the predictive model on multi-site datasets acquired by different protocols/scanners. Specifically, histogram matching is firstly employed before performing tissue segmentation in the iBEAT V2.0 Cloud, then an anatomy prior guided CNN is applied for segmentation/parcellation. The anatomy prior, for example, cortical thickness is within a certain

range, is site-independent and thus makes the iBEAT V2.0 Cloud robust with multi-site infant MRI data. Furthermore, our work ignores specific features of the cortical surface that precisely quantify early brain development, for example, mean curvature, sulcal depth, local gyrification, cortical thickness, surface area, and deep sulcal landmarks, as well as their hemispheric asymmetry and covariance networks. Integration of these surface-based characterizations could further improve the autism predictive performance. Another limitation is that we only explore sex information its relation to ASD in this study, we did not investigate other potentially relevant information, such as intelligence quotients (Yankowitz et al., 2020), ages (Shen et al., 2018), or other non-imaging parameters. As an additional limitation, the proposed method has only been validated on a small number of 24-month-old subjects.

Though we demonstrated that the segmentation and parcellation performed by the iBEAT V2.0 Cloud improved the robustness of the predictive model for multi-site datasets acquired by different protocols/scanners, integration of a data harmonization technique, such as ComBat (Fortin et al., 2017; Fortin et al., 2018), into the iBEAT V2.0 Cloud might further improve its performance. Our future work will be dedicated to further improving our approach by dealing with multi-site data heterogeneity in an end-to-end manner using Generative Adversarial Networks (Goodfellow et al., 2020), which can transform multiple distributions from multi-site datasets to the same distribution for additional benefit in the subsequent processing and learning. Future work will also include integrating more surface-based characterizations, intelligence quotients, ages and other non-imaging parameters into the framework for early-stage status prediction. It would also be useful to explore relationships between the areas highlighted in attention mapping and their correlations with ASD. For example, we observed that the amygdala and hippocampus hold more specific information than other structures for predicting ASD, which indicates that it may be worth exploring the relationship between these two regions relating to autism; such relationships may serve as additional potential markers for early detection of ASD. Future efforts by our group will be targeted toward extending our methods to subjects less than 24 months of age, such as 6 months old, which is a meaningful age from a clinical perspective, and will include more subjects from different scanners/protocols to validate the robustness of the method.

In conclusion, the proposed method can be used to distinguish early-stage ASD and NC subjects with multiple scanners/protocols. With image processing by iBEAT V2.0 Cloud, this method can directly and automatically predict the status of a given subject by using corresponding segmentation and parcellation maps. Compared with previous methods, the proposed method does not rely upon predefined landmarks/biomarkers, which increases its efficiency and effectiveness.

Furthermore, results from attention mapping were consistent with known effects of ASD, thus supporting the prediction ability of the method. While the proposed method has many benefits that suggest high potential for clinical applicability, including enhanced discrimination ability with favorable SPE/SEN/ACC, future refinement efforts may further improve its performance.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS
Kun Gao: Conceptualization, Methodology, Software, and Writing - Original Draft; **Yue Sun:** Software, Validation, Writing - Review & Editing; **Sijie Niu:** Conceptualization, Validation, Writing - Review & Editing; **Li Wang:** Conceptualization, Formal analysis, Resources, Writing - Review & Editing, Supervision, Funding acquisition.

ORCID
Kun Gao https://orcid.org/0000-0003-1729-2700
Li Wang https://orcid.org/0000-0003-2165-0080

REFERENCES
Anello, A., Reichenberg, A., Luo, X., Schmeidler, J., Hollander, E., Smith, C. J., Puleo, M. C., Kryzak, A. L., & Silverman, J. M. (2009). Brief report: Parental age and the sex ratio in autism. *Journal of Autism and Developmental Disorders*, *39*(10), 1487–1492. https://doi.org/10.1007/s10803-009-0755-y

Chen, H., Duan, X., Liu, F., Lu, F., Ma, X., Zhang, Y., Uddin, Q. L., & Chen, H. (2016). Multivtiate classification of autism spectrum disorder using frequency-specific resting-state functional connectivity—A multi-center study. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *64*, 1–9. https://doi.org/10.1016/j.pnpbp.2015.06.014

Chen, Y. C., Chen, C., Martínez, R. M., Fan, Y. T., Liu, C. C., Chen, C. Y., & Cheng, Y. (2021). An amygdala-centered hyperconnectivity signature of threatening face processing predicts anxiety in youths with autism spectrum conditions. *Autism Research*. https://doi.org/10.1002/aur.2595

Chopra, S., Hadsell, R., & LeCun, Y. (2005). *Learning a similarity metric discriminatively, with application to face verification*. Paper presented at the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). https://doi.org/10.1109/CVPR.2005.202

Committee on Educational Interventions for Children with Autism, Board on Behavioral, C, Sensory Sciences, Board on Children,

Youth and Families, Division of Behavioral and Social Sciences and Education, & National Research Council. (2001). *Educating children with autism.* National Academy Press.

Conti, E., Retico, A., Palumbo, L., Spera, G., Bosco, P., Biagi, L., Fiori, S., Tosetti, M., Cipriani, P., Cioni, G., Muratori, F., Chilosi, A., & Calderoni, S. (2020). Autism spectrum disorder and childhood apraxia of speech: Early language-related hallmarks across structural MRI study. *Journal of Personalized Medicine*, *10*(4), 275. https://doi.org/10.3390/jpm10040275

Dillenburger, K. (2014). Why early diagnosis of autism in children is a good thing. *The Conversation.*

Fortin, J.-P., Cullen, N., Sheline, Y. I., Taylor, W. D., Aselcioglu, I., Cook, P. A., Adams, P., Cooper, C., Fava, M., McGrath, P. J., Mclnnis, M., Phillips, M. L., Trivedi, M. H., Weissman, M. M., & Shinohara, R. T. (2018). Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage*, *167*, 104–120. https://doi.org/10.1016/j.neuroimage.2017.11.024

Fortin, J.-P., Parker, D., Tunç, B., Watanabe, T., Elliott, M. A., Ruparel, K., Roalf, D. R., Satterthwaite, T. D., Gur, R. C., Gur, R. E., Schultz, R. T., Verma, R., & Shinohara, R. T. (2017). Harmonization of multi-site diffusion tensor imaging data. *NeuroImage*, *161*, 149–170. https://doi.org/10.1016/j.neuroimage.2017.08.047

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, *63*(11), 139–144. https://doi.org/10.1145/3422622

Hall, D., Huerta, M. F., McAuliffe, M. J., & Farber, G. K. (2012). Sharing heterogeneous data: The national database for autism research. *Neuroinformatics*, *10*(4), 331–339. https://doi.org/10.1007/s12021-012-9151-4

Hazlett, H. C., Gu, H., Munsell, B. C., Kim, S. H., Styner, M., Wolff, J. J., Elison, J. T., Swanson, M. R., Zhu, H., Botteron, K. N., Collins, D. L., Constantino, J. N., Dager, S. R., Estes, A. M., Evans, A. C., Fonov, V. S., Gerig, G., Kostopoulos, P., McKinstry, R. C., … the IBIS Network. (2017). Early brain development in infants at high risk for autism spectrum disorder. *Nature*, *542*(7641), 7348–7351. https://doi.org/10.1038/nature21369

Howell, B. R., Styner, M. A., Gao, W., Yap, P.-T., Wang, L., Baluyot, K., Yacoub, E., Chen, G., Potts, T., Salzwedel, A., Li, G., Gilmore, J. H., Piven, J., Smith, J. K., Shen, D., Ugurbil, K., Zhu, H., Lin, W., & Elison, J. T. (2019). The UNC/UMN baby connectome project (BCP): An overview of the study design and protocol development. *NeuroImage*, *185*, 891–905. https://doi.org/10.1016/j.neuroimage.2018.03.049

Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. (2020). Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(8), 2011–2023. https://doi.org/10.1109/TPAMI.2019.2913372

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1006/nimg.2002.1132

Katuwal, G. J., Baum, S. A., & Michael, A. M. (2018). Early brain imaging can predict autism: Application of machine learning to a clinical imaging archive. *bioRxiv*. https://www.biorxiv.org/content/10.1101/471169v1.abstract

Lai, M.-C., & Szatmari, P. (2020). Sex and gender impacts on the behavioural presentation and recognition of autism. *Current Opinion in Psychiatry*, *33*(2), 117–123. https://doi.org/10.1097/YCO.0000000000000575

Li, G., Liu, M., Sun, Q., Shen, D. & Wang, L. (2018). Early diagnosis of autism disease by multi-channel CNNs. Paper presented at the International Workshop on Machine Learning in Medical Imaging. https://doi.org/10.1007/978-3-030-00919-9_35

Lian, C., Liu, M., Wang, L. & Shen, D. (2019). End-to-end dementia status prediction from brain MRI using multi-task weakly-supervised attention network. Paper presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention. https://doi.org/10.1007/978-3-030-32251-9_18

Liu, M., Lian, C., & Shen, D. (2020). Anatomical-landmark-based deep learning for alzheimer's disease diagnosis with structural magnetic resonance imaging. In *Deep learning in healthcare* (pp. 127–147). Springer. https://doi.org/10.1007/978-3-030-32606-7_8

Maenner, M. J., Shaw, K. A., & Baio, J. (2020). Prevalence of autism spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring network, 11 sites, United States, 2016. *MMWR Surveillance Summaries*, *69*(4), 1–12. https://doi.org/10.15585/mmwr.ss6904a1

Makropoulos, A., Robinson, E. C., Schuh, A., Wright, R., Fitzgibbon, S., Bozek, J., Counsell, S. J., Steinweg, J., Vecchiato, K., & Passerat-Palmbach, J. (2018). The developing human connectome project: A minimal processing pipeline for neonatal cortical surface reconstruction. *NeuroImage*, *173*, 88–112. https://doi.org/10.1016/j.neuroimage.2018.01.054

Mostapha, M., Casanova, M. F., Gimel'farb, G., & El-Baz, A. (2015). Towards non-invasive image-based early diagnosis of autism. Paper presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention. https://doi.org/10.1007/978-3-319-24571-3_20

Mottron, L., Duret, P., Mueller, S., Moore, R. D., d'Arc, B. F., Jacquemont, S., & Xiong, L. (2015). Sex differences in brain plasticity: A new hypothesis for sex ratio bias in autism. *Molecular Autism*, *6*(1), 1–19. https://doi.org/10.1186/s13229-015-0024-1

Nordahl, C. W., Iosif, A.-M., Young, G. S., Hechtman, A., Heath, B., Lee, J. K., Libero, L., Reinhardt, V. P., Winder, B., Amaral, D. G., Rogers, S., Solomon, M., Ozonoff, S. (2020). High psychopathology subgroup in young children with autism: Associations with biological sex and amygdala volume. *Journal of the American Academy of Child & Adolescent Psychiatry*, *59*(12), 1353-1363.e2. https://doi.org/10.1016/j.jaac.2019.11.022

Pierce, K., Gazestani, V. H., Bacon, E., Barnes, C. C., & Courchesne, E. (2019). Evaluation of the diagnostic stability of the early autism spectrum disorder phenotype in the general population starting at 12 months. *JAMA Pediatrics*, *173*(6), 578–587. https://doi.org/10.1001/jamapediatrics.2019.0624

Postorino, V., Fatta, L. M., De Peppo, L., Giovagnoli, G., Armando, M., Vicari, S., & Mazzone, L. (2015). Longitudinal comparison between male and female preschool children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *45*(7), 2046–2055. https://doi.org/10.1007/s10803-015-2366-0

Rogers, T. D., McKimm, E., Dickson, P. E., Goldowitz, D., Blaha, C. D., & Mittleman, G. (2013). Is autism a disease of the cerebellum? An integration of clinical and pre-clinical research. *Frontiers in Systems Neuroscience*, *7*, 15. https://doi.org/10.3389/fnsys.2013.00015

Saeed, F., Eslami, T., Mirjalili, V., Fong, A., & Laird, A. (2019). ASD-DiagNet: A hybrid learning approach for detection of autism spectrum disorder using fMRI data. *Frontiers in Neuroinformatics*, *13*, 70. https://doi.org/10.3389/fninf.2019.00070

Shen, M. D., Nordahl, C. W., Li, D. D., Lee, A., Angkustsiri, K., Emerson, R. W., Rogers, S. J., & Amaral, D. G. (2018). 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*(11), 895–904. https://doi.org/10.1016/S2215-0366(18)30294-3

Shen, M. D., Nordahl, C. W., Young, G. S., Wootton-Gorges, S. L., Lee, A., Liston, S. E., Harrington, K. R., Ozonoff, S., & Amaral, D. G. (2013). Early brain enlargement and elevated extra-axial fluid in infants who develop autism spectrum disorder. *Brain*, *136*(9), 2825–2835. https://doi.org/10.1093/brain/awt166

Shi, F., Yap, P.-T., Wu, G., Jia, H., Gilmore, J. H., Lin, W., & Shen, D. (2011). Infant brain atlases from neonates to 1-and 2-year-olds. *PLoS One*, *6*(4), e18746. https://doi.org/10.1371/journal.pone.0018746

Stoodley, C. J. (2014). Distinct regions of the cerebellum show gray matter decreases in autism, ADHD, and developmental dyslexia. *Frontiers in Systems Neuroscience*, 8, 92. https://doi.org/10.3389/fnsys.2014.00092

Sun, Y., Gao, K., Wu, Z., Li, G., Zong, X., Lei, Z., Wei, Y., Ma, J., Yang, X., Feng, X., Zhao, L., Phan, T., Shin, J., Zhong, T., Zhang, Y., Yu, L., Li, C., Basnet, R., Ahmad, M. O., ... Wang, L. (2021). Multi-site infant brain segmentation algorithms: The iSeg-2019 challenge. *IEEE Transactions on Medical Imaging*, 40(5), 1363–1376. https://doi.org/10.1109/TMI.2021.3055428

Wang, L., Li, G., Shi, F., Cao, X., Lian, C., Nie, D., Liu, M., Zhang, H., Li, G., Wu, Z., Lin, W. & Shen, D. (2018). Volume-based analysis of 6-month-old infant brain MRI for autism bio-marker identification and early diagnosis. Paper presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention. https://doi.org/10.1007/978-3-030-00931-1_47

Wang, L., Nie, D., Li, G., Puybareau, E., Dolz, J., Zhang, Q., Wang, F., Xia, J., Wu, Z., Chen, J., Thung, K.-M., Bui, T., Shin, J., Zeng, G., Zheng, G., Fonov, V. S., Xu, Y., Moeskops, P., Pluim, J. P. W., ... Shen, D. (2019). Benchmark on automatic six-month-old infant brain segmentation algorithms: The iSeg-2017 challenge. *IEEE Transactions on Medical Imaging*, 38(9), 2219–2230. https://doi.org/10.1109/TMI.2019.2901712

Xiao, X., Fang, H., Wu, J., Xiao, C., Xiao, T., Qian, L., Liang, F., Xiao, Z., Chu, K., & Ke, X. (2017). Diagnostic model generated by MRI-derived brain features in toddlers with autism spectrum disorder. *Autism Research*, 10(4), 620–630. https://doi.org/10.1002/aur.1711

Xu, Q., Zuo, C., Liao, S., Long, Y., & Wang, Y. (2020). Abnormal development pattern of the amygdala and hippocampus from youth to adolescent with autism. *Journal of Clinical Neuroscience*, 78, 327–332. https://doi.org/10.1016/j.jocn.2020.03.049

Yale-Medicine. (2021). Functional MRI of the Brain. https://ym.care/mbq

Yankowitz, L. D., Herrington, J. D., Yerys, B. E., Pereira, J. A., Pandey, J., & Schultz, R. T. (2020). Evidence against the "normalization" prediction of the early brain overgrowth hypothesis of autism. *Molecular Autism*, 11(1), 1–17. https://doi.org/10.1186/s13229-020-00353-2

Zhang, H., Shen, D., & Lin, W. (2019). Resting-state functional MRI studies on infant brains: A decade of gap-filling efforts. *NeuroImage*, 185, 664–684. https://doi.org/10.1016/j.neuroimage.2018.07.004

Zhang, J., Gao, Y., Gao, Y., Munsell, B. C., & Shen, D. (2016). Detecting anatomical landmarks for fast Alzheimer's disease diagnosis. *IEEE Transactions on Medical Imaging*, 35(12), 2524–2533. https://doi.org/10.1109/TMI.2016.2582386

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.