

# Identifying Children with Autism Spectrum Disorder Based on Their Face Processing Abnormality: A Machine Learning Framework

Wenbo Liu, Ming Li, and Li Yi

The atypical face scanning patterns in individuals with Autism Spectrum Disorder (ASD) has been repeatedly discovered by previous research. The present study examined whether their face scanning patterns could be potentially useful to identify children with ASD by adopting the machine learning algorithm for the classification purpose. Particularly, we applied the machine learning method to analyze an eye movement dataset from a face recognition task [Yi et al., 2016], to classify children with and without ASD. We evaluated the performance of our model in terms of its accuracy, sensitivity, and specificity of classifying ASD. Results indicated promising evidence for applying the machine learning algorithm based on the face scanning patterns to identify children with ASD, with a maximum classification accuracy of 88.51%. Nevertheless, our study is still preliminary with some constraints that may apply in the clinical practice. Future research should shed light on further valuation of our method and contribute to the development of a multitask and multimodel approach to aid the process of early detection and diagnosis of ASD. *Autism Res* 2016, 00: 000–000. © 2016 International Society for Autism Research, Wiley Periodicals, Inc.

**Keywords:** autism spectrum disorder; face processing; eye tracking; machine learning

## Introduction

Autism Spectrum Disorder (ASD) is a heritable, lifelong neurodevelopmental disorder with complicated causes and courses [Amaral, Schumann, & Nordahl, 2008]. One of the key symptoms in ASD is their impaired social interaction and interpersonal communication [DSM-V, APA, 2013]. Since interpersonal interaction and communication rely heavily on interpreting facial cues of other people, research on the face processing in ASD has attracted intensive attention in the last decade [see Tanaka & Sung, 2013, for a review]. Overall, individuals with ASD have difficulty recognizing human faces and interpreting facial emotions [Tanaka & Sung, 2013]. Eye tracking studies further reveal that individuals with ASD exhibit reduced to attention to human faces, especially the eye region, relative to typically developing (TD) individuals [see Falck-Ytter & von Hofsten, 2011, for a review]. Tanaka and Sung [2013] attributed this eye-avoidance-face-processing pattern of ASD to their adaptive strategy to avoid the social threatening and discomfort elicited by direct eye contact. The reduced attention to faces and diminished eye contact has been found in individuals with ASD as early as the

first year of life [e.g. Osterling, Dawson, & Munson, 2002; Zwaigenbaum, Bryson, Rogers, Roberts, Brian, & Szatmari, 2005]. Jones and Klin [2013] found that a decline in eye fixation at 2–6 months could predict later diagnosis of ASD. These findings suggest the possibility of using eye movements during face processing as a potential indicator of ASD. In this paper, we explored the possibility of using the machine learning algorithm to identify ASD based on their eye movements when viewing faces.

Machine learning is a procedure that trains the computer algorithm to analyze a set of observed data and statistically learns the latent patterns. It has been used to perform a variety of prediction tasks in psychology recently, for example, the emerging field of multimodel human sensing, where the states of human, such as emotion, are analyzed with computer vision and speech techniques based on machine learning [Bartlett, Littlewort, Lainscsek, Fasel, & Movellan, 2004]. Machine learning has also been applied to autism research in several previous studies based on the behavioral observations or brain activities [e.g. Crippa et al., 2015; Deshpande, Libero, Sreenivasan, Deshpande, & Kana, 2013; Duda, Kosmicki, & Wall, 2014; Kosmicki, Sochat, Duda,

From the From the Sun Yat-sen University Carnegie Mellon University Joint Institute of Engineering, Sun Yat-sen University, Guangzhou Higher Education Mega Center, Guangzhou, China (W.L., M.L.); Sun Yat-sen University Carnegie Mellon University Shunde International Joint Research Institute, Shunde, Guangdong, China (M.L.); Department of ECE, Carnegie Mellon University, Pittsburgh, PA, USA (W.L.); Department of Psychology and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, China (L.Y.)

Received November 02, 2015; accepted for publication January 30, 2016

Address for correspondence and reprints: Li Yi, Department of Psychology and Beijing Key Laboratory of Behavior and Mental Health, Peking University, 5 Yiheyuan Road, Beijing 100871, China. E-mail: yilipku@pku.edu.cn; Ming Li, Sun Yat-sen University Carnegie Mellon University Joint Institute of Engineering, Sun Yat-sen University, No 132 East Waihuan Road, Guangzhou higher education mega center, Guangzhou, 510006, P.R. China. Email: liming46@mail.sysu.edu.cn

Published online 00 Month 2016 in Wiley Online Library (wileyonlinelibrary.com)

DOI: 10.1002/aur.1615

© 2016 International Society for Autism Research, Wiley Periodicals, Inc.

**Table 1. Participant Characteristics in Each Group**

	Male/Female	Mean age in years (SD)	NVIQ <sup>a</sup> raw score (SD)	Autism quotient (SD)	
	ASD	25/4	7.90 (1.45)	22.29 (10.80)	85.48 (14.90)
	TD-Age	25/4	7.86 (1.38)	29.90 (9.96)	62.74 (12.57)
	TD-Ability	25/4	5.74 (1.01)	22.28 (7.90)	64.03 (12.21)
Difference ( <i>t</i> -test)	ASD vs. TD-Age	N/A	0.09	-2.77**	6.15***
	ASD vs. TD-Ability	N/A	6.56***	0.00	6.00***
	TD-Age vs. TD-Ability	N/A	6.69***	3.23**	-0.39

Note. <sup>a</sup>IQ was measured by the Combined Raven Test (CRT-C2).

\*\*\**p* < .001; \*\**p* < .01; \**p* < .05

& Wall, 2015; Stahl, Pickles, Elsabbagh, Johnson, & BASIS Team, 2012; Zhou, Yu, & Duong, 2014]. Despite such a fact, the importance and power of machine learning is yet to be fully discovered in autism research. While machine learning in general usually consists of feature extraction, feature selection, model learning, and prediction, a number of machine-learning-based autism studies mainly focus on how to effectively select a subset of features from large amounts of existing features of the standardized diagnostic scales to shorten the diagnosis time. For instance, Kosmicki et al. [2015] proposed a machine-learning-based feature selection framework to reduce the number of behavioral features and measurements in Autism Diagnostic Observation Schedule (ADOS). Duda et al. [2014] used machine learning to train a classifier which can reduce 72% in length in the ADOS-G test with 97% accuracy. Some recent autism studies start to address the task of classification and prediction besides the feature selection [Crippa et al., 2015; Stahl et al., 2012; Zhou et al., 2014]. For example, a machine learning framework was proposed to classify low-functioning children with ASD based on the kinematic analysis of a reach-to-drop task [Crippa et al., 2015]. This motivated us to take the face scanning patterns, which have been consistently found to be atypical in ASD by previous studies, as input to classify children with ASD. Eye movements encode rich information about the attention distribution and cognitive strategies during face viewing that may indicate the potential risk of ASD, such as the fixation durations and counts at different facial areas, the speed and direction of the saccades, as well as the temporal information of the face scanning pattern. Automatically handling eye gaze data with machine learning methods makes the prediction process much more scalable than manually doing so. The fundamental purpose of this paper is to propose a machine learning framework which learns from the observed face scanning patterns to automatically identify children with ASD. We hope that such a framework can generate useful midlevel features in the ASD evaluation, and that by adopting eye movement, subjective factors can be reduced to make the ASD evaluation a more objective process.

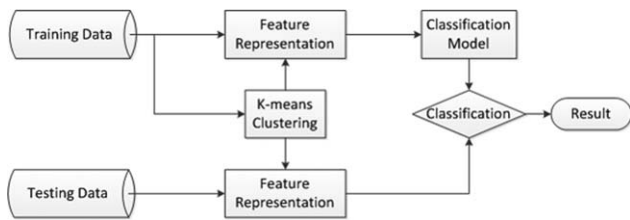
Our work to some extent shares the similar spirit with Crippa et al. [2015], except that we focus on analyzing face scanning patterns to predict ASD. We adopted the eye movement dataset from a previous published work, which asked children with and without ASD to recognize same- and other-race face while their eye movements being tracked [Yi, Quinn, Fan, Huang, Feng, Li, & Lee, 2016]. Different from the previous literature focusing on the statistical significance of ASD symptoms conveyed by the face scanning patterns, we addressed the prediction problem and sought to propose a machine learning solution to measure the potential ASD risk based on the face scanning patterns in a face recognition task. Particularly, we used a data-driven approach to extract features from the face scanning data and a support vector machine (SVM) to do the classification. The predictive value of this machine learning model was evaluated in terms of its accuracy, specificity, and sensitivity, and so on.

## Method

### Description of the Dataset

The dataset used in the current paper included three groups of participants: 29 4- to 11-year-old Chinese children with ASD, 29 Chinese TD children matched with the chronological age, and another group of 29 Chinese TD children matched with IQ (see Table 1 for details). All children with ASD were diagnosed by experienced clinicians and met the diagnostic criteria for Autism Spectrum Disorder according to the DSM-IV [APA, 2000]. Due to the limited access to the ADI-R and ADOS in China, we confirmed the diagnosis using the Chinese version of Autism Spectrum Quotient: Children's Version [AQ-Child; Auyeung, Baron-Cohen, Wheelwright, & Allison, 2008].

Children were asked to memorize six faces (three Chinese faces as the same-race faces and three Caucasian faces as the other-race faces), and later tested to recognize these faces from 18 novel faces, including same- and other-race faces (width: 500 pixels, height: 700 pixels, resolution: 72 pixels per in.). All face stimuli were gray-scale and front-view, with their external



**Figure 1.** Flowchart of the proposed ASD classification framework.

features (e.g. hair and ears) removed with an ellipse-shaped window. There were three study blocks, and in each study block, children were asked to remember two faces (one Chinese face and one Caucasian face, each presented for 3 sec). Each study block was followed by three test blocks in which children were asked to identify whether each face was seen before or not. Each test block comprised two target faces and two foil faces, which were presented until children responded. Children’s eye movements during the study and test phases were recorded by a Tobii T60 eye-tracker (sample rate: 60 Hz; both eyes were tracked) with the Tobii Studio software. More details of the participants, the material, and the experimental procedures were provided in the Yi et al. [2016].

Yi et al. [2016] analyzed the eye movement data based on the area of interest (AOI) approach to compare the fixation duration between groups within each predefined face region (e.g. eyes, nose, and mouth). In this paper, we proposed a new framework based on the machine learning algorithm to analyze the eye movement data during face processing so as to identify the ASD symptoms.

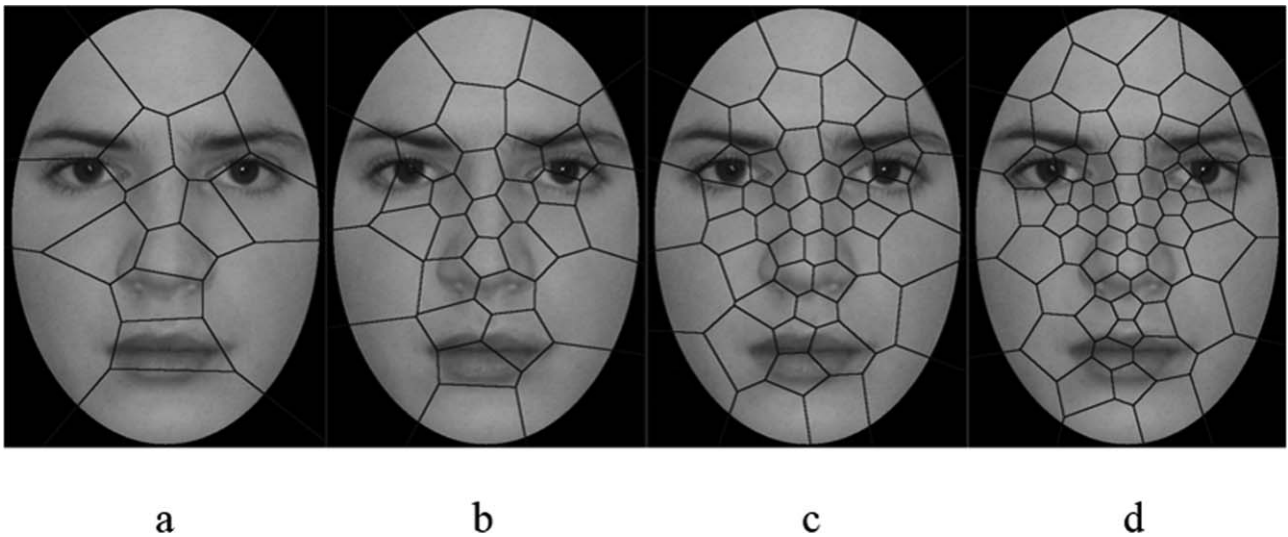
**Framework of Proposed Method.** We proposed a new framework using the machine learning algorithm to identify children with ASD based on the face scanning patterns. Particularly, we obtained the observations of eye movement patterns from categorized participants (ASD or TD children), with which we attempted to design a system to automatically classify children with ASD on a new observation. In the rest of this section, we will describe the detailed procedures of our proposed ASD classification framework, including the feature representation to select relevant features, and the classification to assign group membership based on the selected features (Figure 1). We performed the whole machine learning process on the Matlab platform.

**Feature Representation.** Feature representation is a crucial part of our classification framework to select relevant features for the classification purpose. A discrimi-

native (good) feature should maximally reveal the statistical difference between participants from different groups, while being minimally sensitive to intra-group variations. While the sequence of eye fixation coordinates or face regions can be incorporated as a temporal feature, we did not adopt these temporal features here due to the sparseness of cross-region transition in our training data. We therefore considered “orderless” features where the measurement is not sensitive to temporal order of coordinates. What we measured was the frequency distribution of coordinates which treated all face scanning coordinates equally without temporal information. More importantly, we used the frequency distribution as a discriminative feature for the ASD classification, considering the existing evidence on the correlation between the coordinate frequency distribution when scanning faces and the ASD symptoms [Yi et al., 2013, 2014]. Numerous studies have indicated that children and adults with ASD show atypical visual attentions to faces compared to their TD counterparts [e.g. Pelphrey, Sasson, Reznick, Paul, Goldman, & Piven, 2002; Yi et al., 2013]. Such a face scanning atypicality was directly reflected in the abnormality of ASD in the distributions of fixation coordinates, which serves as a feature in our framework. The feature representation includes two procedures: the facial region partitioning with *K*-means and the histogram feature extraction.

We performed the quantization of fixation coordinates with the *K*-means algorithm, where fixation coordinates are clustered and divided into *K* different clusters with distinct cluster centroids, as shown in Figure 2. The *K*-means quantization was conducted based on the fixation coordinate data of all participants. Each observed coordinate was assigned to the cluster with the closest centroid. Such quantization results in the partitioning of face images into *K* different cell-like regions, such that fixations falling into the same region indicate close proximity in the visual attention location [Hartigan & Wong, 1979]. Compared to the well-known Area of Interest (AOI) based approach, our quantization strategy was more data-driven oriented. The AOI approach is a “top-down” process which determines the partitioned region boundary empirically and could be influenced by the semantic meaning of face parsing without statistical justification [Yi et al., 2014]. In contrast, our data-driven approach can represent face scanning “hot spots” by generating partitions based on statistical distribution of coordinates.

Given the sequence of the fixation coordinates from each face viewed by every participant, we assigned each coordinate to the most proximal cluster centroid obtained by *K*-means and counted the number of assignments for each cluster. Then the assignment counts were normalized by being divided by the total number of coordinates. As a result, a histogram feature



**Figure 2.** Illustration of partitioned face regions by  $K$ -means with different cluster numbers ( $K$ ). (a)  $K = 16$ . (b)  $K = 32$ . (c)  $K = 48$ . (d)  $K = 64$ .

was defined to decode the frequency distribution of the visual attentions on each part of the face.

Since the extracted histogram is an image-level feature encoding the visual attention on a single face, we repeated the histogram extraction process for every face viewed by each participant, to obtain a training set whose labels are determined by the participant categories (i.e. ASD group, aged-matched TD group, and IQ-matched TD group).

**Classification.** The classification is the process to use the selected features to assign group labels to the participants. Given the labeled features, the classification algorithm could build a classifier that assigns new examples into different categories. The classification process included the following steps: the generation of the training and testing data, the image-level classification, and the participant-level classification.

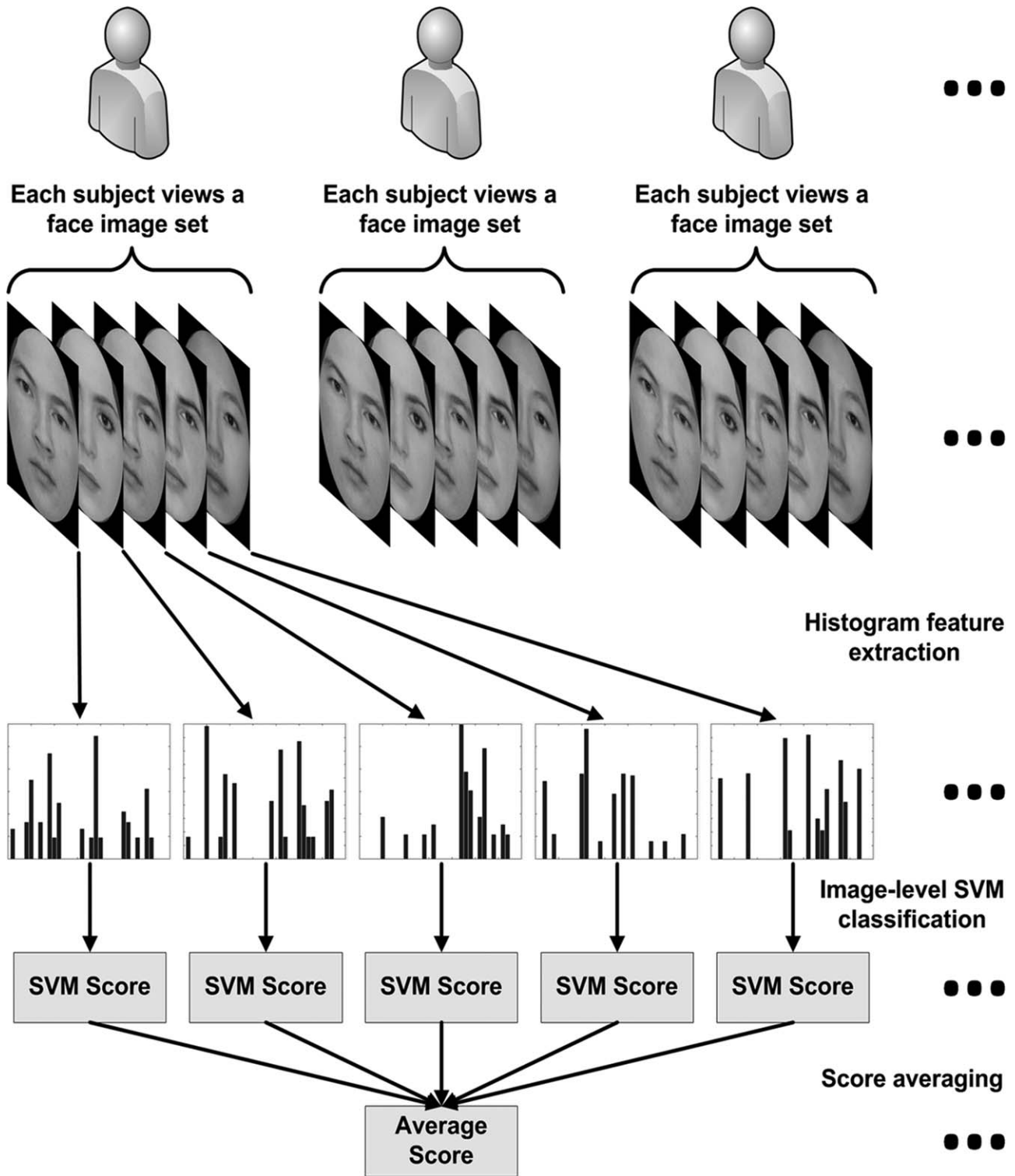
Given a set of labeled participants, we used the “leave-one-out” cross-validation strategy to separate the original data into the training dataset and the testing dataset [Vapnik, 1998]. Each time, one out of all participants was selected sequentially as the testing participant while the classification model was learned according to the histogram features from the rest of the participants. The learned model was then tested, and then both the image-level scores and the participant-level scores for the test participant were returned. Such a “leave-out-and-evaluate” process was repeatedly performed for all participants included in our dataset.

We started with the image-level ASD classification based on the extracted histograms containing the visual attention information on single faces, followed by the participant-level classification. At the training stage, a

SVM classifier was trained based on the labeled histograms [Cortes & Vapnik, 1995; Chapelle, Haffner, & Vapnik, 1999]. The SVM attempted to find a linear decision boundary with a maximum margin separating the data into two classes. Considering that the data were not linearly separable, we adopted the Radial Basis Function (RBF) kernel SVM which performed a nonlinear projection of the data into a high-dimensional space to make the data more linearly separable [Hsu, 2003].

During the testing phase, the learned SVM model made a classification for the group membership using each testing histogram feature with a corresponding confidence score. The sign of the score can be either positive or negative to indicate the classification of the histogram feature. The absolute value of the confidence score measures the distance of the testing sample from the decision boundary. A higher confidence score indicates a more confident classification decision.

The image-level ASD classification indicates the likelihood of the ASD symptoms only based on the face scanning patterns from a single face. It was more meaningful to make the classification decision at the participant level to indicate the likelihood of the ASD symptoms for each participant. We therefore defined the participant-level classification score as the average image-level classification score of each participant, as shown in Figure 3. Considering that the imbalanced ASD and TD training set sizes may cause biased SVM classifications, we introduced a flexible threshold instead of zero to determine the final classification labels. The participants with the classification score above the threshold were labeled as individuals potentially with ASD, while those ones whose classification score below the threshold were labeled as TD individuals.



**Figure 3.** Participant-level classification based on image-level information fusion.

**Data Analysis.** In order to evaluate the performance of our proposed classification framework, we computed the accuracy, specificity, sensitivity, the Receiver Operator Characteristic (ROC) curves, and the corresponding area under the curve (AUC) of our framework. The

accuracy was calculated as the count of the correctly predicted participants divided by the count of all participants. The sensitivity (the true positive rate, TPR) measured the percentage of participants with ASD correctly classified as ASD by the proposed framework. The

**Table 2. The Accuracy, AUC, Sensitivity, and Specificity of the Proposed Framework**

	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)
<b>All Face Images</b>	<b>88.51***</b>	<b>93.10</b>	<b>86.21</b>	<b>89.63</b>
Same-Race Faces	81.61	75.86	84.48	82.40
Other-Race Faces	90.80	96.55	87.93	94.41
<b>ASD vs. both TD Groups</b>	<b>88.51</b>	<b>93.10</b>	<b>86.21</b>	<b>89.63</b>
ASD vs. TD-Age Groups	84.48	89.66	79.31	85.37
ASD vs. TD-IQ Groups	86.21	89.66	82.76	88.94

\*\*\* $p < .001$

specificity (the true negative rate, TNR) measured the percentage of the TD participants correctly predicted as TD individuals by our framework.

We then grouped the participants into four categories according to their ground truth (actual positive or negative) and the predicted group membership (predicted positive or negative). Here positive refers to the ASD group, while negative refers to the TD group, with the TD-age and the TD-IQ groups combined. Based on the counts of four categories, we computed the accuracy, the sensitivity, the specificity, and the false positive rate (FPR) according to the following formulas:

$$Accuracy = \frac{TP + TN}{N}$$

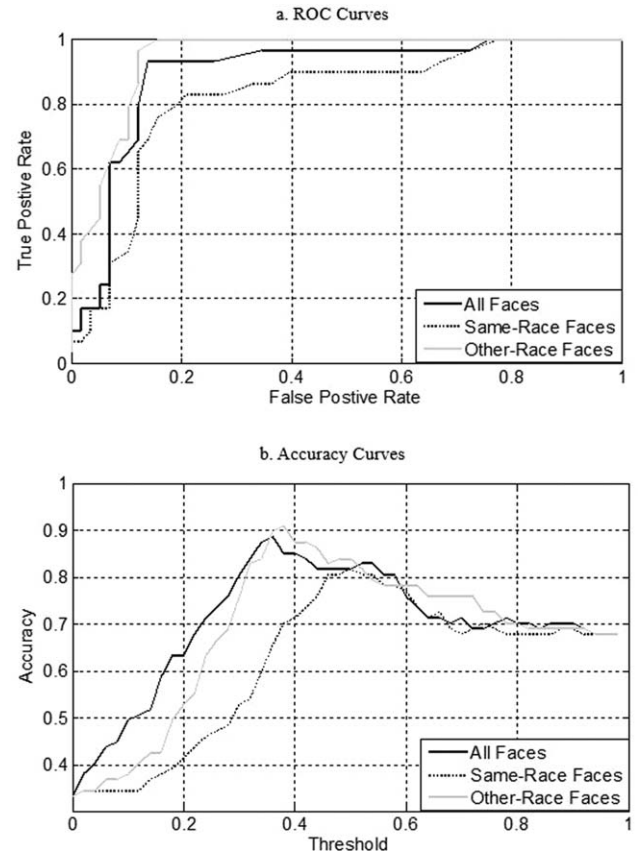
$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$FPR = \frac{FP}{FP + TN}$$

Then, a global threshold was set for all participants according to their predicted scores to determine their predicted group membership. Different thresholds were evaluated to determine different accuracy, sensitivity and specificity values. We then computed the ROC curves by plotting all the TPR vs. FPR pairs, and the AUC using the Matlab platform.

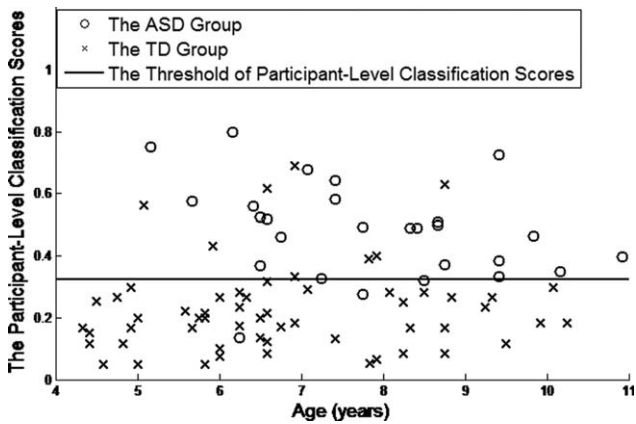
In order to identify which face regions best discriminated between groups, we further conducted additional data analysis using a simple feature selection method. First, we computed two mean histogram features from the ASD group (denoted as  $Feat_{ASD}$ ) and TD group (denoted as  $Feat_{TD}$ ) respectively by taking the mean of all image-wise histogram features with the same group labels (ASD/TD). We then computed the mean feature difference using the following formula:  $Feat_{Dif} = Feat_{ASD} - Feat_{TD}$ , where bins in  $Feat_{Dif}$  with larger absolute values indicated the regions that were more discriminative. Regions with positive values were preferred by the ASD group, whereas regions with negative values indicated the preference for the TD group. After computing  $Feat_{Dif}$ , we could choose highest bins and show the visualized result of selected discriminative regions.



**Figure 4.** The ROC curves (a) and the accuracy curves (b) of all faces, same-race and other-race faces.

## Results

In this section, we comprehensively reported the performance of our proposed framework on the datasets. First, we reported the performance of our proposed method (i.e. accuracy, sensitivity, specificity, ROC, and AUC curves) for all face images, and also for the same-race and the other-race faces separately (Table 2 and Figure 4). Second, the participant-level classification scores were plotted against participants' chronological age in Figure 5. We also reported the performance of our framework for the ASD group from the TD-age and the TD-IQ groups, respectively (Table 2).



**Figure 5.** The relationship between the participant-level classification scores and the age of participants in years, with an optimal decision threshold of 0.3255.

To evaluate the performance of our proposed framework, we respectively computed the accuracy curves and the ROC curves for all faces, the same-race faces, and the other-race faces, respectively, as shown in Figure 4. The cluster numbers were tuned for different experiments to reach the best accuracy. The cluster numbers were, respectively, set to 16, 64, and 96 for the same-race, other-race, and all faces in the current model. The accuracy, sensitivity, specificity, AUC for all faces and for same- and other-race faces were summarized in Table 2. For all faces, the maximum classification accuracy reached 88.51% (sensitivity = 93.10%, specificity = 86.21%), and the AUC reached 89.63%. A better performance of the framework based on the other-race faces (accuracy = 90.80%, sensitivity = 96.55%, specificity = 87.93%, & AUC = 94.41%) in comparison with the same-race faces (accuracy = 81.61%, sensitivity = 75.86%, specificity = 84.48%, and AUC = 82.40%) was observed.

We then plotted the classification scores against the chronological age of each participant in Figure 5, with  $x$ -axis corresponding to the chronological age of participants and  $y$ -axis standing for the participant-level classification scores. The circle markers and the cross markers represent the ASD and the TD groups, respectively. The horizontal line represents the optimal decision threshold (0.3255),<sup>1</sup> at which the classification accuracy is maximized. As shown in Figure 5, vast majority of children in the ASD group scored above the threshold, while majority of children in the TD group scored below the threshold, demonstrating that the classifications made by the proposed framework were mostly robust and unambiguous. No effect of age was observed

<sup>1</sup>A participant is predicted as having ASD if the participant-level score is higher than this threshold, and not having ASD otherwise.

for the classification performance of the framework ( $r = .17$ ,  $p = .11$ , Pearson Correlation).

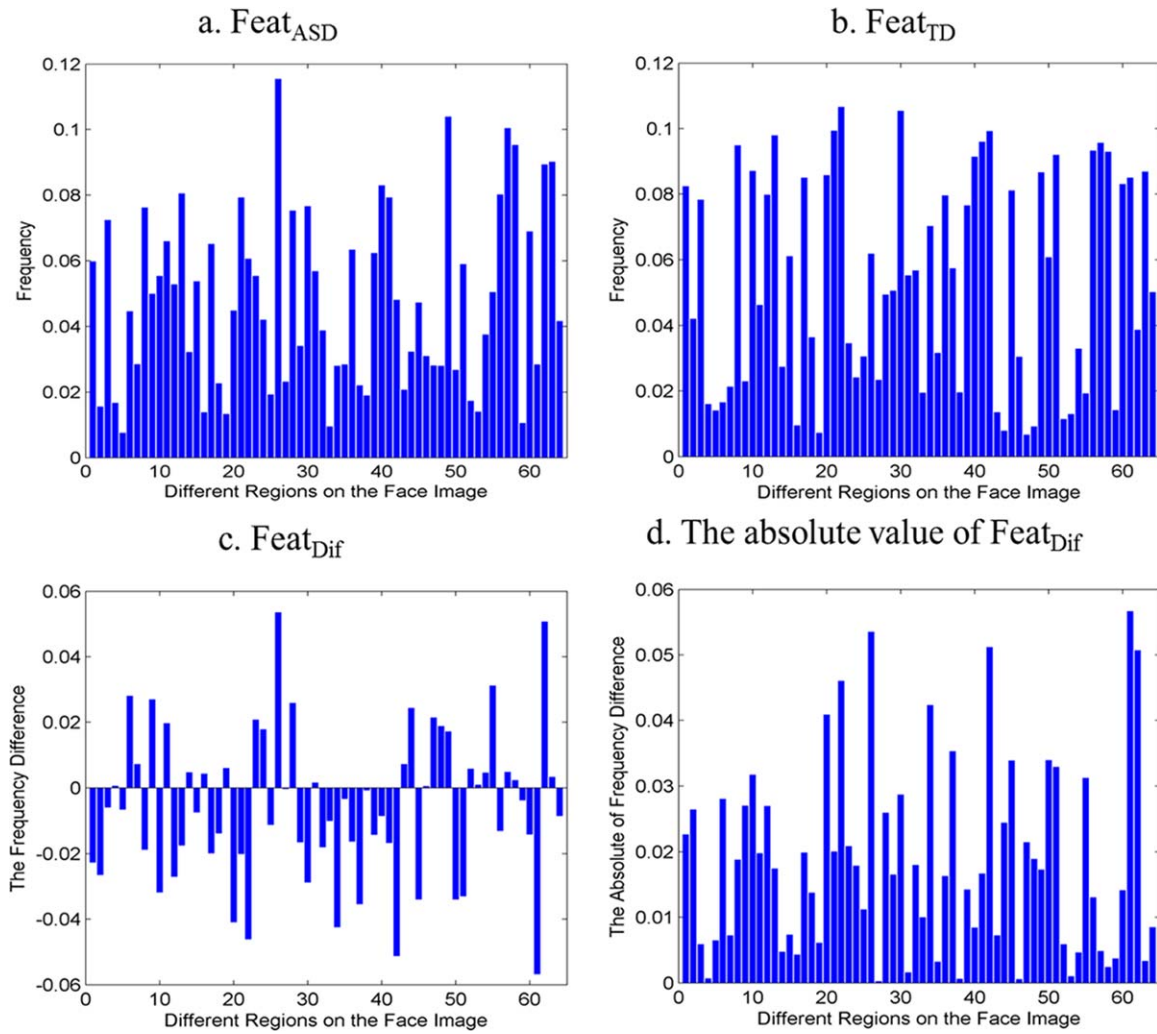
Besides calculating the accuracy of the classification framework, we also evaluated the performance of our algorithm to discriminate the ASD group from the TD-age and the TD-IQ groups separately. The maximum accuracy and the AUC of the framework were computed, respectively, for the ASD vs. TD-age comparison and the ASD vs. TD-IQ comparison. The results, as listed in Table 2, showed that for the ASD vs. TD-age comparison, the maximum accuracy of reached 84.48% (with the sensitivity = 89.66%, specificity = 79.31%) and the AUC reached 85.37%; for the ASD vs. TD-IQ comparison, the maximum accuracy reached 86.21% (with the sensitivity = 89.66%, specificity = 82.76%) and the AUC reached 88.94%. The results indicated that the proposed framework can discriminatively classify ASD participants from both TD groups.

After computing the accuracy of our proposed data analysis method, we also analyzed the contribution of different face regions to the classification. Figure 6a and b show  $Feat_{ASD}$  and  $Feat_{TD}$ , respectively: the  $x$ -axis indicates the 64 regions obtained by  $K$ -means on the face image, and the  $y$ -axis corresponds to the frequency of fixation points on each region. Figure 6c and d, respectively, shows the values of  $Feat_{Dif}$  and its absolute values. The most discriminative regions derived by  $Feat_{Dif}$  are highlighted in Figure 7: purple regions indicated the most discriminative areas where the ASD group spent longer time looking at (the right eye and the area slightly above the mouth), while cyan regions indicated the most discriminative area where the TD group spent longer time looking at (the area slightly below the left eye), given the same observation time.

We then computed the histogram features on the selected discriminative regions with our proposed method, and used the selected features to re-perform the prediction with the SVM framework proposed in the paper. Interestingly, we found that the top seven discriminating dimensions resulted in the accuracy of 79.31%, the sensitivity of 68.96%, and the specificity of 84.48%. This meant that if we only select the seven most discriminative features from the original 64 features, we could still achieve a reasonably good performance.

## Discussion

The current paper proposed a machine learning framework to identify children with ASD based on the face scanning eye movement patterns. We adopted a data-driven feature extraction method and a SVM to do the classification. Results indicated that our machine learning model could deliver rather good performance of



**Figure 6.** The mean histogram features of the ASD (a) and the TD (b) groups, and their differences (c and d).

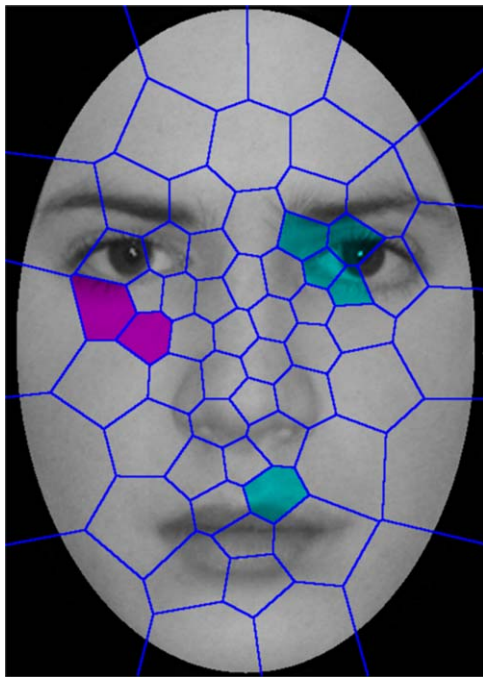
classifying the ASD and TD groups based on the face scanning patterns (accuracy = 88.51%; specificity = 86.21; sensitivity = 93.10%; AUC = 89.63%). In a word, our findings evidently manifest the effectiveness and feasibility of applying the machine learning algorithm based on the face scanning patterns in classifying and predicting ASD.

Besides the overall performance evaluation of our model, we have also identified the most discriminative facial areas that could result in 79.31% of accuracy to classify the ASD and the TD groups. Particularly, the TD group looked longer than the ASD group in the right eye (from the observer’s view), and the area slightly above the month; the ASD group looked longer than the TD group at the area slightly below the left eye (from the observer’s view). That is, these facial areas are the most efficient ones in distinguishing the face scanning patterns of ASD and TD groups. This finding may also imply the abnormal face scanning patterns in individuals with ASD, consistent with the previous litera-

ture: they exhibit reduced attention to main facial features (eyes, nose, and mouth), compared to the typical population. [e.g. Chawarska & Shic, 2009; Dalton et al., 2005; Pelphrey et al. 2002; Spezio, Adolphs, Hurley, & Piven, 2007; see Falck-Ytter & von Hofsten, for a review]. The longer looking time at the area below the eye in ASD was also found in a previous study using a data-driven analytic approach [Yi et al., 2013]. This novel looking pattern in ASD, as argued in Yi et al. [2013], may result from ASD children’s strong tendency to avoid direct eye contact with another person, proposed as the “eye-avoidance” hypothesis in autism face processing [Tanaka & Sung, 2013].

Overall, our study provides promising findings and implications for the possibility of early detection and a computer-aid diagnostic approach of ASD, which create big challenges for the current clinical practice of ASD. Despite its early onset, ASD is usually diagnosed several years later, mainly based on interviews with parents and clinical behavioral observations (i.e. ADOS, ADI-R).





**Figure 7.** The seven most discriminative regions derived from  $Feat_{Dif}$  (Purple regions indicated discriminative the ones preferred by ASD, cyan regions indicated the ones preferred by TD).

Recognizing early signs of ASD provides opportunity for early detection and early intervention. The early signs and potential predictors of ASD included deficits in joint attention, pretend play, perspective taking, responses to own name, imitation, and so on [Lai, Lombardo, & Baron-Cohen, 2013]. These early signs of ASD, however, are not reliably observed in infants, which makes the early detection of ASD very challenging. Meanwhile, since the accuracy of the diagnosis relies heavily on the clinicians' expertise and experience, a computer-aid diagnosis will improve the condition by providing a more objective and reliable approach. Our study is one of the first attempts to address these challenges. Moreover, modern remote eye tracking device allows the participants to move their head freely and its high sampling rates make it possible to collect massive eye movement data in a short time. All these technique advances, combined with well-developed data analytic and classification algorithm, will facilitate the progress of developing a reliable approach for early detection and computer-aid diagnosis of ASD.

However, in spite of our promising findings and their potential application prospect, there is still a long way to go before we could apply this procedure in the actual clinical practice due to several constraints. First, cautions should be taken when using the face scanning pattern as a biomarker for ASD diagnosis: the interpretation of our method is constrained by the prevalence of the ASD in the general population (around 1%),

which is much lower than the rate of ASD in our sample (33%) [Yerys, B. E., & Pennington]. Second, the face scanning patterns could be age and culture adapted, since people at different ages and in different cultures may scan face differently [e.g. Fu, Hu, Wang, Quinn, & Lee, 2012; Liu, Quinn, Wheeler, Xiao, Ge, & Lee, 2011; Wheeler, Anzures, Quinn, Pascalis, Omrin, & Lee, 2011; Yi et al., 2016]. Also, other characteristics of the ASD sample (e.g. severity of the symptom, cognitive function, etc.) may also affect their face scanning patterns [e.g. Yi et al., 2014]. Therefore, these factors should be considered when applying our model to classify children with ASD for particular cases. Third, considering our sample size in the current study is relatively small for the purpose of pattern classification, our method and conclusions should be replicated in future studies with larger samples in order to validate our machine learning approach. Fourth, the predictive value of our model should be evaluated in the future studies to track a group of high-risk infants for several years, to determine whether the face processing pattern in infancy could actually predict their severity of symptom when being diagnosed several years later. Fifth, the current study used a data-driven approach to extract features from the eye tracking data. The algorithm could be improved with more sophisticated models to determine which and how many (at least) features of eye movement indices could be identified to be use in the clinical practice to identify children with ASD. Sixth, since the atypical face scanning pattern is only one of numerous potential indicator of ASD symptoms, future investigations should be undertaken to apply our method to different stimuli (social or nonsocial stimuli), and tasks. For example, Gliga and colleagues [2015] found that the enhanced visual search performance of infants at 9 months old, reflected in their eye movements, could predict the severity of their autism symptoms at 15 and 24 months. This is another piece of promising evidence of using eye movement indices as an early indicator to predict later autism symptoms. Finally, our measure of the face scanning eye movement patterns should be combined with other types of psychological and physiological measures (e.g. brain activities, skin conductance, speech, motions, body gestures, and facial expressions, etc.) to obtain a more comprehensive multimodel measure of the risk of ASD to aid the process of diagnosis and early detection for ASD.

In summary, despite of the above limitations, our study provides preliminary but promising evidence for using the machine learning algorithm based on the face scanning patterns to predict the ASD phenotype. This could be one of the first attempts to develop a computer-aid early detection and diagnosis system to support the clinical practice of the screening and diagnosis of ASD. Future research should focus on further

evaluating and validating the model in multiple ways, and generalizing our method into different populations, stimuli, tasks, and measures.

## Acknowledgments

The authors are grateful to Dr. Sheng Li, Dr. Xueqin Wang, Dr. Bhiksha Raj, Zhiding Yu, Pengli Li, Jiao Li, for their generous assistance in completing the study. The authors declare no conflict of interest. Grant sponsor: National Natural Science Foundation of China; Grant numbers: 31571135, 31200779, 61401524. Grant sponsor: National Science Foundation of Guangdong Province; Grant number: 2014A030313123. Grant sponsor: Fundamental Research Funds for the Central Universities; Grant number: 15lgjc12, 15lgjc40. Grant sponsor: Guangdong Shunde SYSU-CMU Joint Research Institute; Grant number: 20140302; Grant sponsor: CMU-SYSU Collaborative Innovation Research Center.

## References

- Amaral, D. G., Schumann, C. M., & Nordahl, C. W. (2008). Neuroanatomy of autism. *Trends in Neurosciences*, 31(3), 137–145.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.) (text revision). Washington, DC: American Psychiatric Press.
- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Press.
- Auyeung, B., Baron-Cohen, S., Wheelwright, S., & Allison, C. (2008). The Autism Spectrum Quotient: Children's Version (AQ-Child). *Journal of Autism and Developmental Disorders*, 38, 1230–1240.
- Bartlett, M. S., Littlewort, G., Lainscsek, C., Fasel, I., & Movellan, J. (2004). Machine learning methods for fully automatic recognition of facial expressions and facial actions. In *IEEE International Conference on Systems, Man and Cybernetics* (vol. 1, pp. 592–597). Netherlands: The Hague.
- Chapelle, O., Haffner, P., & Vapnik, V. N. (1999). Support vector machines for histogram-based image classification. *IEEE Transactions on Neural Networks*, 10, 1055–1064.
- Chawarska, K., & Shic, F. (2009). Looking but not seeing: Atypical visual scanning and recognition of faces in 2 and 4-year-old children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 39(12), 1663–1672.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273–297.
- Crippa, A., Salvatore, C., Perego, P., Forti, S., Nobile, M., Molteni, M., et al. (2015). Use of machine learning to identify children with autism and their motor abnormalities. *Journal of Autism and Developmental Disorders*, 45, 1–11.
- Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., et al. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nature Neuroscience*, 8(4), 519–526.
- Deshpande, G., Libero, L. E., Sreenivasan, K. R., Deshpande, H. D., & Kana, R. K. (2013). Identification of neural connectivity signatures of autism using machine learning. *Frontiers in Human Neuroscience*, 7, 670.
- Duda, M., Kosmicki, J. A., & Wall, D. P. (2014). Testing the accuracy of an observation-based classifier for rapid detection of autism risk. *Translational Psychiatry*, 4, e424.
- Falck-Ytter, T., & von Hofsten, C. (2011). How special is social looking in ASD: A review. *Progress in Brain Research*, 189, 209–222.
- Fu, G., Hu, C. S., Wang, Q., Quinn, P. C., & Lee, K. (2012). Adults scan own- and other- race faces differently. *PLoS One*, 7, e37688.
- Gliga, T., Bedford, R., Charman, T., Johnson, M. H., & BASIS Team. (2015). Enhanced visual search in infancy predicts emerging autism symptoms. *Current Biology*, 25(13), 1727–1730.
- Hartigan, J. A., & Wong, M. A. (1979). Algorithm as 136: A k-means clustering algorithm. *Applied Statistics*, 28(1), 100–108.
- Hsu, C. W., Chang, C. C., & Lin, C. J. (2003). A practical guide to support vector classification, available at: <https://www.cs.sfu.ca/people/Faculty/teaching/726/spring11/svmguide.pdf>, 1–16.
- Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. *Nature*, 504(7480), 427–431.
- Kosmicki, J. A., Sochat, V., Duda, M., & Wall, D. P. (2015). Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. *Translational Psychiatry*, 5, e514.
- Lai, M. C., Lombardo, M. V., & Baron-Cohen, S. (2013). Autism. *The Lancet*, 383(9920), 896–910.
- Liu, S., Quinn, P. C., Wheeler, A., Xiao, N., Ge, L., & Lee, K. (2011). Similarity and difference in the processing of same- and other-race faces as revealed by eye tracking in 4- to 9-month-olds. *Journal of Experimental Child Psychology*, 108, 180–189.
- Osterling, J. A., Dawson, G., & Munson, J. A. (2002). Early recognition of 1-year-old infants with autism spectrum disorder versus mental retardation. *Development and Psychopathology*, 14, 239–251.
- Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of Autism and Developmental Disorders*, 32(4), 249–261.
- Spezio, M. L., Adolphs, R., Hurley, R. S. E., & Piven, J. (2007). Abnormal use of facial information in high-functioning autism. *Journal of Autism and Developmental Disorders*, 37(5), 929–939.
- Stahl, D., Pickles, A., Elsabbagh, M., Johnson, M. H., & BASIS Team. (2012). Novel machine learning methods for ERP analysis: A validation from research on infants at risk for autism. *Developmental Neuropsychology*, 37, 274–298.
- Tanaka, J. W., & Sung, A. (2013). The “eye avoidance” hypothesis of autism face processing. *Journal of Autism and Developmental Disorders*, doi: 10.1007/s10803-013-1976-7.
- Vapnik, V. (1998). *Statistical Learning Theory*. New York: Wiley-Interscience.

- Wheeler, A., Anzures, G., Quinn, P. C., Pascalis, O., Omrin, D. S., & Lee, K. (2011). Caucasian infants scan own- and other-race faces differently. *PLoS One*, 6(4), e18621.
- Yerys, B. E., & Pennington, B. F. (2011). How do we establish a biological marker for a behaviorally defined disorder? Autism as a test case. *Autism Research*, 4(4), 239–241.
- Yi, L., Fan, Y., Quinn, P. C., Feng, C., Huang, D., Li, J., et al. (2013). Abnormality in face scanning by children with autism spectrum disorder is limited to the eye region: Evidence from multi-method analyses of eye tracking data. *Journal of Vision*, 13, 1–13.
- Yi, L., Feng, C., Quinn, P. C., Ding, H., Li, J., Liu, Y., et al. (2014). Do Individuals with and without Autism Spectrum Disorder Scan Faces Differently? A New Multi-Method Look at an Existing Controversy. *Autism Research*, 7, 72–83.
- Yi, L., Quinn, P. C., Fan, Y., Huang, D., Feng, C., Li, J., & Lee, K. (2016). Children with Autism Spectrum Disorder scan own-race faces differently than other-race faces. *Journal of Experimental Child Psychology*, 141, 177–186.
- Zhou, Y., Yu, F., & Duong, T. (2014). Multiparametric MRI characterization and prediction in autism spectrum disorder using graph theory and machine learning. *PloS One*, 9, e90405–e90405.
- Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., & Szatmari, P. (2005). Behavioral manifestation of autism in the first year of life. *International Journal of Developmental Neuroscience*, 23, 143–152.