



Stratification of Children with Autism Spectrum Disorder Through Fusion of Temporal Information in Eye-gaze Scan-Paths

25

ADHAM ATYABI, University of Colorado Colorado Springs

FREDERICK SHIC, University of Washington and Seattle Children's Research Institute

JIAJUN JIANG, Old Dominion University

CLAIRE E. FOSTER, Binghamton University (SUNY), Binghamton

ERIN BARNEY, Seattle Children's Research Institute

MINAH KIM, University of Virginia

BEIBIN LI, University of Washington

PAMELA VENTOLA, Yale University

CHUNG HAO CHEN, Old Dominion University

Background: Looking pattern differences are shown to separate individuals with Autism Spectrum Disorder (ASD) and Typically Developing (TD) controls. Recent studies have shown that, in children with ASD, these patterns change with intellectual and social impairments, suggesting that patterns of social attention provide indices of clinically meaningful variation in ASD.

Method: We conducted a naturalistic study of children with ASD ($n = 55$) and typical development (TD, $n = 32$). A battery of eye-tracking video stimuli was used in the study, including Activity Monitoring (AM), Social Referencing (SR), Theory of Mind (ToM), and Dyadic Bid (DB) tasks. This work reports on the feasibility of spatial and spatiotemporal scanpaths generated from eye-gaze patterns of these paradigms in stratifying ASD and TD groups.

Algorithm: This article presents an approach for automatically identifying clinically meaningful information contained within the raw eye-tracking data of children with ASD and TD. The proposed mechanism utilizes combinations of eye-gaze scan-paths (spatial information), fused with temporal information and pupil velocity data and Convolutional Neural Network (CNN) for stratification of diagnosis (ASD or TD).

Results: Spatial eye-gaze representations in the form of scanpaths in stratifying ASD and TD (ASD vs. TD: DNN: 74.4%) are feasible. These spatial eye-gaze features, e.g., scan-paths, are shown to be sensitive to factors

This study was funded by: NIH K01 MH104739 and Simons Foundation #383661.

Authors' addresses: A. Atyabi, University of colorado colorado springs, 1420 Austin Bluffs Pkwy, Colorado Springs, CO, USA 80918; email: aatyabi@uccs.edu; F. Shic, Seattle Children's Research Institute/University of Washington School of Medicine, 1920 Terry Ave M/S Cure-3, Seattle, WA 98101; email: fshic@uw.edu; J. Jiang, Old Dominion University, 231 Kaufman Hall, Norfolk, VA 23529; email: jjian001@odu.edu; C. E. Foster, Binghamton University – State University of New York, 4400 Vestal Parkway East, Binghamton, NY 13902; email: cfoster7@binghamton.edu; E. Barney, Seattle Children's Research Institute/University of Washington School of Medicine, 1920 Terry Ave M/S Cure-3, Seattle, WA 98101; email: erin.barney.consulting@gmail.com; M. Kim, 510 7 1/2th St SW, Charlottesville VA 22903; email: mk7kc@virginia.edu; B. Li, Microsoft Research, 14820 NE 36th St, Redmond, WA 98052; email: beibin.li@microsoft.edu; P. Ventola, Yale Child Study Center, 230 South Frontage Rd., New Haven, CT 06520; email: pamela.ventola@yale.edu; C. H. Chen, 231 Kaufman hall, old dominion university, Norfolk va 23529; email: CXCHEN@odu.edu.

Updated author affiliation: Beibin Li, Microsoft Research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2023 Association for Computing Machinery.

1556-4681/2023/02-ART25 \$15.00

<https://doi.org/10.1145/3539226>

mediating heterogeneity in ASD: age (ASD: 2–4 y/old vs. 10–17 y/old CNN: 80.5%), gender (Male vs. Female ASD: DNN: 78.0%) and the mixture of age and gender (5–9 y/old Male vs. 5–9 y/old Female ASD: DNN: 98.8%). Limiting scan-path representations temporally increased variance in stratification performance, attesting to the importance of the temporal dimension of eye-gaze data. Spatio-Temporal scan-paths that incorporate velocity of eye movement in their images of eye-gaze are shown to outperform other feature representation methods achieving classification accuracy of 80.25%.

Conclusion: The results indicate the feasibility of scan-path images to stratify ASD and TD diagnosis in children of varying ages and gender. Infusion of temporal information and velocity data improves the classification performance of our deep learning models. Such novel velocity fused spatio-temporal scan-path features are shown to be able to capture eye gaze patterns that reflect age, gender, and the mixed effect of age and gender, factors that are associated with heterogeneity in ASD and difficulty in identifying robust biomarkers for ASD.

CCS Concepts: • Computing methodologies → Supervised learning by classification; Neural networks;
• Applied computing → Health Informatics;

Additional Key Words and Phrases: Autism Spectrum Disorder, eye tracking, eye-gaze scan-path, Convolution Neural Network

ACM Reference format:

Adham Atyabi, Frederick Shic, Jiajun Jiang, Claire E. Foster, Erin Barney, Minah Kim, Beibin Li, Pamela Ventola, and Chung Hao Chen. 2023. Stratification of Children with Autism Spectrum Disorder Through Fusion of Temporal Information in Eye-gaze Scan-Paths. *ACM Trans. Knowl. Discov. Data.* 17, 2, Article 25 (February 2023), 20 pages.

<https://doi.org/10.1145/3539226>

1 INTRODUCTION

Autism is a neurodevelopmental disorder associated with social communication deficits and repetitive behaviors [12]. Assessments performed by trained psychologists and clinicians and caregiver reports are the main sources of clinical and research assessment of ASD. In addition to being subjective, these assessments provide minimal information about the underlying mechanism of the disorder. Adding to these limitations the observed heterogeneity in autism, the need for a robust approach that has less reliance on expert clinical assessments and can also address the heterogeneity of autism disorder is clear and of immediate interest in the autism research community.

Deficit in social attention is a known symptom of autism and one of the focus areas in autism biomarker discovery [13]. This atypicality in social attention in ASD is observed in a variety of research studies and experimental modalities [14–16]. **Eye-tracking (ET)** is shown to be able to assess social attention presented as video clips. This is because ET is able to provide moment-to-moment and frame-by-frame evaluation of how and when social (e.g., faces) and non-social (e.g., toys, photo frames on the wall, and background) components of the scenes are looked at. Over the last decades, an extensive body of research has been performed focusing on the intersection between autism disorder and social attention deficit, revealing differences in looking at faces and other social information [17, 18]. Being non-invasive, safe, and tolerated within a wide age range of participants, from infancy to adulthood, makes ET a suitable approach for studying adaptive and cognitive functioning [19, 34].

Automatic detection of autism diagnosis based on eye-gaze information attracted attention with several studies utilizing various representations of eye-gaze data combined with machine learning methods to predict patients' diagnosis [41, 42]. Deep neural network is considered state-of-the-art in machine learning and has been successfully incorporated into computer vision studies. Deep visual attention networks are found effective for predicting gaze location [2–6].

Carette et al. [1] used saccades information and **long short-term memory (LSTM)** network to predict autism diagnosis with 83% accuracy using data from 32 children aged between 8 and

10. Li et al. [42] introduced **Oculomotor Behavior Framework (OBF)** model that is capable of learning OBs from unsupervised and semi-supervised tasks. Using a dataset of 49 children (38 ASD) authors achieved 80% classification accuracy in stratifying ASD from TD. Elbattah et al. [8] studied the utility of deep autoencoder for identifying clusters of ASD and non-ASD using scan-path patterns of 59 children (mean age of 7.88 years old). Considering 2 and 3 clusters, Elbattah et al., showed evidence of ASD heterogeneity in the sense of inclusion of ASD participants in all clusters ranging from 28% to 94% contribution to clusters. Pusiol, et al. [9], studied feasibility of vision-based gender-specific stratification of **Fragile X Syndrome (FXS)**, a case of autism with a genetic cause, and **developmental disorder (DD)**. In their study, eye gaze data of 70 participants were used in combination with a modified LSTM and showed that it is feasible to stratify male-FXS and female-FXS gaze patterns from DD with 86% and 91% classification accuracies, respectively. The results reported in this study indicate clear differences in eye gaze patterns of DD, and male and female F XS individuals. However, such stratification analysis between Male vs. female F XS, F XS vs. DD, and F XS vs. TD (Typically developing) is not reported making it difficult to attest to the ability of the proposed approach to address autism heterogeneity. Tao and Shyu [10] in their 2019 Saliency4ASD grand challenge submission proposed SP-ASDNet, a hybrid **Convolutional Neural Network (CNN)** and LSTM network that utilizes eye-gaze scan-path images to diagnose ASD. Tao and Shyu used eye-gaze data from 28 children (5–12 years old) looking at 300 images and were able to achieve 74.22% accuracy on the validation set, although their performance on the testing set was reduced to 55.66% due to over-fitting problem. Wu et al. [11] proposed image-based and synthetic saccade methods that use scan-path images for automatic classification of ASD. The authors used two deep networks with 8 and 10 layers to train the two models they proposed. 2019 Saliency4ASD grand challenge dataset containing scan-paths of 28 children was used and 65.41% and 55.13% classification accuracy is achieved on validation and testing sets, respectively. Li et al. [41] introduced **Sparsely Grouped Input Variables for Neural Network (SGIN)**, a mechanism for automated selection of ET experimental stimuli where high between group discrimination (ASD vs. non-ASD) is observed and regression with clinical variables are achieved.

In this article, four well-studied ET tasks are used for data collection. These tasks we selected considering (a) strong construct performance, (b) between-group discrimination, and (c) relation to ASD symptoms in prior research studies of school-aged children. These tasks are used to record eye-gaze patterns during observation of (1) videos of two adults playing with toys (**Activity Monitoring (AM)** task); (2) videos of three adults having conversation (**Dyadic Bid (DB)**), (3) videos of an adult secretly changing the location of an object placed by another adult (**Theory of Mind (ToM)**), and (4) videos of an adult performing a stressful action (**Social Referencing (SR)**). The objective of this study is to investigate the feasibility of using scan-paths, an eye-gaze spatial information representation mechanism, to predict autism diagnosis.

What are the contributions of this work?

This study introduces the novel ideas of (a) incorporating temporal information to scan-path images developed based on eye-gaze information and (b) infusing eye-gaze velocity in spatiotemporal scan-path samples aiming to improve the informativeness of such gaze data representation and increasing their collective ability to stratify children with Autism Spectrum Disorder from their Typically Developing counter-parts.

2 METHODS AND MATERIALS

2.1 Participant Characteristics

Ninety-nine participants are enrolled in the study, out of which eye gaze information from 86 participants (ASD n = 54, TD n = 32) with valid data are considered. The remaining 86 participants are further grouped into five categories based on their ages. These subgroups (see Table 1) include

Table 1. Participants are Grouped By Age

IF ASD	Dataset No. of Subjects					
	0–1	2–4	5–9	10–17	18+	Total
ASD	3	11	23	13	4	54
TD	5	6	16	5	0	32
Total	8	17	39	18	4	86

Table 2. Participants are Grouped By Gender and Age

	Sex	Male				
		0–1	2–4	5–9	10–17	18+
ASD	Age	0–1	2–4	5–9	10–17	18+
	No. Sub	1	8	19	12	4
	Total	44				
TD	Age	0–1	2–4	5–9	10–17	18+
	No. Sub	1	1	10	3	0
	Total	15				
	Sex	Female				
		0–1	2–4	5–9	10–17	18+
ASD	Age	0–1	2–4	5–9	10–17	18+
	No. Sub	2	3	4	2	0
	Total	10				
TD	Age	0–1	2–4	5–9	10–17	18+
	No. Sub	4	5	6	2	0
	Total	17				

0 to 1 year (0–1), 2 to 4 years (2–4), 5 to 9 years (5–9), 10 to 17 years (10–17) and 18+ years old (18+).

Each subject has a different number of recorded ET videos and the data is unbalanced with respect to diagnosis, age, and gender (see Tables 1 and 2).

3 EXPERIMENT DESIGN

3.1 Activity Monitoring (AM)

Monitoring the activities of others is compromised in children with ASD, [30, 31] with recent studies showing atypical looking patterns during AM probes are present even in adults with ASD. In this task, children are presented with multiple AM trials as developed in [30] and other work. During video trials, the actresses spoke using child-friendly language and directed their eyes to each other (mutual gaze) or the activity regions in between them. The scenes in these videos also contain multiple clusters of age-appropriate and colorful toys that are irrelevant to the activity and are used as distractors. An example of an AM video scene is presented in Figure 1.

3.2 Dyadic Bid Sensitivity

Recent studies have shown that children with ASD look less at the faces of people trying to engage their attention through eye contact and using child-directed speech (i.e., bids for dyadic engagement) [32, 33, 35] as compared to controls. In this task, we extend our previous work examining this phenomenon in toddlers with ASD [32, 33, 36] to a new, more complex, and challenging environment for older children. In this task, multiple actors are seen engaging in conversation/interaction with one another. Periodically, one of the actors looks directly at the camera, “speaking” to the viewer. During the DB, the two actors who were not speaking performed one of two behaviors:



Fig. 1. AM example video scene.



Fig. 2. DB example video scenes.

(1) looked silently at the viewer along with the actor making the bid, or (2) gazed down and made subtle hand movements. Behavior (1) was included to act as both a control (to ensure participants were not looking at the actor making the bid solely because he/she was looking at the participant) as well as a distractor (another oriented face competing for the participant's attention). Behavior (2) was included to act as a foil (a face that is not oriented toward the participant, looking down during a conversation is not a social norm, and the subtle hand movements served as a control for the speaking actor's mouth movements) to make sure participants are not looking at the actor making the bid solely because of the salience of the movement of his/her mouth. This task examines sensitivity to overtures for social engagement by others, a requisite for speech to respond to inquiries and reciprocal conversation. Figure 2 presents an instance of a DB video scene.

3.3 Social Referencing

SR, the process of seeking clarifying context from the faces of others regarding uncertain situations, develops throughout the first year of life [37, 38] but can be deficient in older children with ASD [39]. To examine this social-information-seeking process, we filmed multiple episodes in which an actor engaged in stressful activities (e.g., stacking a tall thin block tower, inflating a balloon till near-burst). These episodes depict the activities' escalation (e.g., balloon continues to expand), critical event (e.g., balloon pops and the pieces begin flapping wildly), and resolution (e.g., actors sigh with relief). This task examines the automaticity of information-seeking from others, a requisite for speech to inquire that is related to sharing behaviors in a speech to comment, and monitor nonverbal information. An example of an SR video scene is presented in Figure 3.

3.4 Theory of Mind (ToM)

Elegantly demonstrated by Senju and colleagues [40] the ability to automatically process others' beliefs as they translate into intentional actions is deficient even in adults with ASD. In this task, a fluid, naturalistic ToM task is created in which a protagonist continuously engages in search/play activities with specific objects while a second, antagonist actor, unbeknownst to the protagonist,

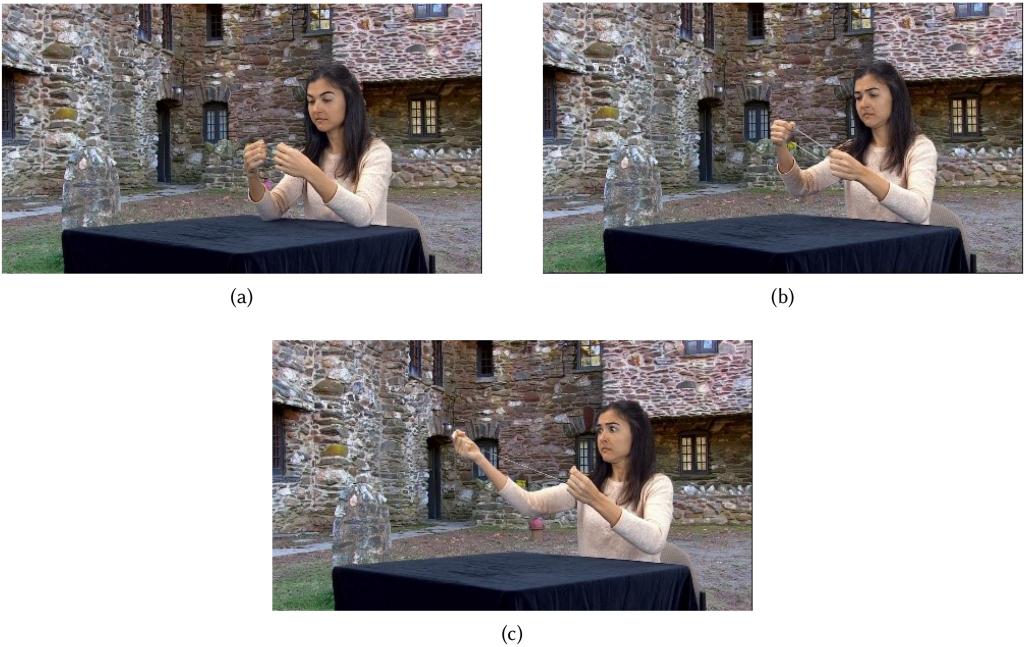


Fig. 3. SR example video scenes.



Fig. 4. TOM example video scenes.

constantly and randomly interferes with the protagonist's goals. For example, the protagonist, wanting to make breakfast, places a box of cereal in a cupboard and then leaves the scene to retrieve a bowl. The antagonist enters, taking the cereal from the cupboard and placing it in a different one. The protagonist comes back and goes to the original cupboard to retrieve the cereal, only to find the cereal has been moved. This task gauges spontaneous interest in others' intentions as well as comprehension of the mental frame and perspective of others and is relevant for perspective-taking and associated skills. An example of a ToM video scene is presented in Figure 4.

The considered experimental paradigm in this study share similarities with recent studies of **autism biomarker Consortium for Clinical Trials (ABC-CT)** [43].

4 COMPUTATIONAL MODELING OF EYE-GAZE SCAN-PATHS

In this study, eye gaze scan-path of participants watching video clips representing AM, DB Sensitivity , SR, and ToM are used. The main objective of this study is to develop flexible machine

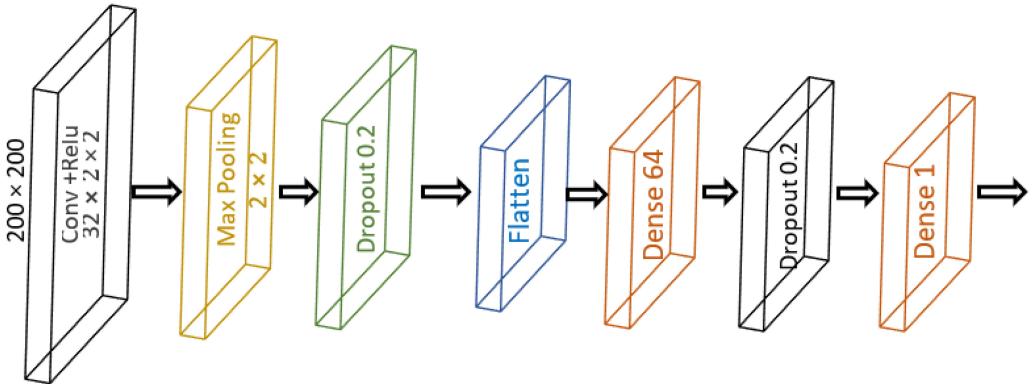


Fig. 5. CNN architecture layout.

learning approaches for parsing heterogeneity within ASD and segregating individuals with ASD from those without. The study utilized CNN for this purpose. The structural layout of CNN used in this study is presented in Figure 5. Several evaluations are performed focusing on general factors such as

- Feasibility of spatial representation of eye-gaze (Scan- Paths) as features for stratifying ASD and TD.
- Factoring ASD heterogeneity (e.g., Age, Gender, and their mixed effects) in stratification of ASD through scan- paths.
- Spatio-Temporal Analysis of eye-gaze data.
- Fusion of gaze velocity in Spatio-Temporal scan-paths and its impact on ASD stratification.

4.1 Spatial Analysis I: Basic Feasibility Evaluation of Scan-paths

4.1.1 Stratification of ASD and TD. The method used in this study is inspired by [22–24]. In this method, the ET scan-paths are generated on a black background for each visualized video clip by each participant. The size of the background is set to $1,680 \times 1,050$ and later the images are resized to 100×100 pixels and changed to grayscale. This process resulted in 1,012 scan-path images in the ASD and 380 in the TD diagnostics categories. **Dense Neural Network (DNN)** and CNN models are utilized. The ratio of training images and testing images is set to 7:3.

It is noticeable that there is a gap in performance between training and testing accuracy and AUC performances. Compared with the CNN model, the DNN model achieved higher accuracy, but it performed poorly in AUC in comparison to CNN.

In order to clarify the identified issue with AUC results, a further analysis is performed using DNN and the dataset provided in [22–24]. This dataset contains 219 images for ASD and 328 images for non-ASD. The results are shown in Table 4. It is noteworthy that the authors indicated that the AUC results can be improved to 0.8120 after augmenting the number of images [22–24].

Comparing the results presented in Tables 3 and 4 impact on AUC estimates are observed with the understanding that the differences between testing accuracy and AUC results using DNN are set to 6% and 7% (dropout 0.5 and 0.2, respectively) in this dataset (see Table 3). Considering results with the database provided in [22–24], the observed differences between testing accuracy and AUC is 3% and 8% with 0.5 and 0.2 dropouts, respectively. The lower AUC results can be explained by having a smaller number of TD training samples compared to ASD. This hinders the network training. One possible resolution to address this issue is to use data augmentation to increase the number of training samples.

Table 3. Results of ASD Classification Used Different Models and Parameter

Models	50 epochs		
	Train ACC	Test ACC	AUC
DNN	0.9969	0.7033	0.6751
Dropout: 0.5	1.0000	0.7297	0.6858
CNN	Train ACC	Test ACC	AUC
Dropout: 0.5	0.9579	0.7249	0.6796
Dropout: 0.2	0.9969	0.7033	0.6614
Models	100 epochs		
	Train ACC	Test ACC	AUC
DNN	1.0000	0.7440	0.6828
Dropout: 0.5	1.0000	0.7440	0.6699
CNN	Train ACC	Test ACC	AUC
Dropout: 0.5	0.9784	0.7249	0.7207
Dropout: 0.2	0.9959	0.7153	0.6889

Number of scan-path images used from ASD and TD after augmentation are 1,012 and 979, respectively.

The bold results are the best performances with DNN and CNN.

Table 4. ASD vs. TD Classification Using Dataset Provided in [22–24]

Models	100 epochs			Dropout
	Train ACC	Test ACC	AUC	
DNN	0.9969	0.7033	0.6751	Dropout: 0.5
	1.0000	0.7697	0.7925	Dropout: 0.2

Numbers of scan-path images used from ASD and non-ASD are 219 and 328, respectively.

4.1.2 Age Classification (2–4 vs. 10–17). Heterogeneity in autism is known to play a significant role in difficulty to identify reliable biomarkers for this disorder. Main factors contributing to autism heterogeneity include genetic variability, comorbidity, and gender [25].

ASD prevalence is found to have a gender bias with one in four individuals with ASD diagnosis being male [26, 27].

ASD heterogeneity and proven difficulty to identify generalizable markers of ASD have led to the presumption of the presence of multiple etiologies rather than a single disorder [28]. The quest to develop personalized medicine for ASD is impacted by ASD heterogeneity [25].

In this experiment, ASD data is used to evaluate the impact of age on stratification of ASD. Two age groups of 2 to 4 years (2–4) and 10 to 17 years (10–17) are considered in this analysis. 2–4 groups contain 11 ASD participants with 199 scan-path sample images. 10–17 group contains 13 participants with 255 scan-path images. Aiming to increase the number of samples, these images are augmented resulting in 871 and 1196 images, respectively. The results are presented in Table 5.

The results indicate high classification accuracy in predicting the age group of ASD participants based on their eye-gaze scan-path images. This in turn indicates that the eye-gaze trajectories of kids with autism spectrum disorder are highly impacted by age.

4.2 Spatial Analysis II: Assessing the Performance Variations Impacted by the Video Clips Observed

This experiment investigates possible effects in scan-path patterns caused by variations in video clips watched by participants. First, total number of times each video clip is watched by participants

Table 5. Results of Age (2–4 vs. 10–17) Classification
Used Different Models and Parameters

Models	50 epochs			
	DNN	Train ACC	Test ACC	AUC
Dropout: 0.2	0.9969	0.7033	0.6751	
Dropout: 0.5	0.9876	0.7665	0.8195	
CNN	Train ACC	Test ACC	AUC	
	Dropout: 0.2	1.0000	0.7955	0.8502
Dropout: 0.5	0.9969	0.8052	0.8682	
Models				
DNN	100 epochs			
	Train ACC	Test ACC	AUC	
Dropout: 0.2	1.0000	0.7552	0.8228	
Dropout: 0.5	1.0000	0.7987	0.8240	
CNN	Train ACC	Test ACC	AUC	
	Dropout: 0.2	0.9993	0.7810	0.8617
Dropout: 0.5	1.0000	0.7987	0.8240	

Number of scan-path images used from 2–4-ASD and 10–17-ASD after augmentation are 871 and 1,196, respectively.

The bold results are the best performances with DNN and CNN.

is counted (see Table 6), and later, to better understand the impact of scan-paths of each video on the overall ability to stratify ASD and TD, these scan-paths are eliminated from the dataset and the process of training DNN and CNN models and their evaluations are repeated.

As presented in Table 6, AM_A4_S6_B4_GM_D1_F1 and sr04 are the two video clips with the highest number of overall views used to verify their impact on feasibility by participants of the study, both having more than 90 views. In this analysis, the scan-path images of these videos are used to stratify ASD and TD categories. To do so, the scan-paths generated from these two video clips are removed from the dataset and the stratification capability of DNN and CNN models are reevaluated. The results presented in Table 7 indicate that in the absence of scan-path samples of AM_A4_S6_B4_GM_D1_F1 video clips, the accuracy of validation is slightly improved compared to the scenario where scan-path images generated from sr04 video clip are removed. This indicates that sr04 video clip is relatively more powerful than AM_A4_S6_B4_GM_D1_F1 in generating scan-paths that can distinguish ASD from TD.

To better assess the impact of various videoclips and their contribution to the observed stratification power of scan-paths, the procedure discussed earlier further considers the second set of videos that were viewed most (e.g., videos that were viewed 60 to 62 times). The results, reported in Table 8, indicate that scan-path images generated from eye-gaze information of participants viewing video clips tom0301A32 and db0101B12 have the highest level of impact in the observed stratification power of DNN. The results also indicate scan-path images generated from eye-gaze information of participants viewing sr01 video clip have the least contribution to the performance of DNN. This is due to achieving the highest classification performance when the scan-path images generated from sr01 are removed from the dataset.

The findings of this experiment indicate that there is a degree of difference between the contribution of each video clip to the overall stratification of ASD and TD.

4.2.1 Gender Classification for ASD Subjects in Trial Level. To understand the impact of sex and age in children with autism on the eye-gaze scan-paths and the feasibility of using these scan-path images in stratifying ASD participants based on their gender, two set of experiments are performed:

Table 6. Total Number of Views of Each Video Used in the Study By Participant

Video Name	Num. of times viewed
Activity Monitoring (AM) video clips	
AM_A0_S3_B3_GA_D1_F0	61
AM_A0_S6_B1_GA_D1_F1	30
AM_A1_S2_B0_GM_D1_F0	30
AM_A1_S5_B2_GA_D1_F1	1
AM_A1_S7_B7_GA_D1_F1	1
AM_A2_S7_B2_GA_D1_F1	61
AM_A4_S6_B4_GM_D1_F1	90
AM_A7_S4_B6_GM_D1_F1	1
AM_A7_S6_B3_GA_D1_F1	29
Dyadic Bids (DB) video clips	
db0101B12	60
db0201A21	60
db0301B33	60
db0401C22	60
db0501C13	60
db0601A31	61
db1303B11	1
db1403A32	1
db1503A23	1
db1603B22	1
db1703C31	1
db1803C13	1
Social Referencing (SR) video clips	
sr01	62
sr02f	61
sr03f	61
sr04	92
sr06	31
sr07	31
sr09f	1
sr10	1
sr11f	1
sr12	1
sr15f	30
Theory of Mind (ToM) video clips	
tom0101A31	61
tom0201A12	61
tom0301A32	61
tom0401A11	60
tom0501A31	61
tom0601A12	61
tom1303B11	1
tom1403B32	1
tom1503B11	1
tom1603B12	1
tom1703B31	1
tom1803B32	1
Total Number	1,414

Number of scan-path images used from ASD and TD after augmentation are 975 and 362, respectively.

Table 7. Contribution of Scan-path Images of Most Viewed (90–92 Times View) Video Clips on Stratification Power of DNN and CNN

Model	Video Clip Removed from dataset	Training Accuracy	Test Accuracy	AUC
DNN 75% Train 25% Test	AM_A4_S6_B4_GM_D1_F1	100%	77.61%	0.6882
	sr04	100%	73.01%	0.7121
CNN 75% Train 25% Test	AM_A4_S6_B4_GM_D1_F1	100%	74.54%	0.7244
	sr04	98.46%	73.62%	0.7386

Number of scan-path images used from ASD and TD after augmentation when AM_A4_S6_B4_GM_D1_F1 video is removed are 954 and 350, respectively, and after sr04 is removed are 953 and 348, respectively.

Table 8. Contribution of Scan-path Images of Second Most Viewed (60–2 Times View) Video Clips on Stratification Power of DNN

DNN model, 75% Train, 25% Test, 100 epochs			
Video Clip Removed from dataset	Training Accuracy	Test Accuracy	AUC
AM_A0_S3_B3_GA_D1_F0	100	72.87	64.72
AM_A2_S7_B2_GA_D1_F1	100	74.78	67.09
sr01	100	74.98	69.25
sr02f	100	73.93	65.51
sr03f	100	74.07	67.17
tom0101A31	100	73.93	65.92
tom0201A12	100	74.49	67.28
tom0301A32	100	70.21	64.54
tom0501A31	100	74.55	65.56
tom0601A12	100	71.02	63.78
tom0401A11	100	72.35	63.50
db0601A31	100	72.53	64.22
db0101B12	100	70.76	66.16
db0201A21	100	74.75	68.36
db0301B33	100	74.09	64.54
db0401C22	100	74.46	68.59
db0501C13	100	74.24	67.56

Bold results indicate the best performance achieved by each set of videos, representing different type of stimuli such as social referencing (sr) theory of mind (Tom) and diadic bid (db).

(1.1) Gender classification in all ASD subjects: The results reported in Table 9 indicate validation accuracy of 0.7804% reflecting feasibility of using scan-paths for stratifying male and female children with autism. This in turn attest to considerable differences in scan-paths of female and male children with ASD enrolled in the study.

(1.2) Gender classification in 5 to 9 years old ASD subjects: Considering the low number of female ASD participants in the study, the group with the largest number of female ASD participants, 5 to 9 years old, is considered in this analysis. The low number of female ASD participants results in a much lower number of available training scan-path samples which in turn negatively impacts DNN's training. To circumvent this problem, three degrees of sample augmentations are considered aiming to increase the number of scan-path samples. A high classification accuracy value among male and female ASD participants is indicative of substantial difference among these two subcategories of ASD in response to an ET stimulus. The results

Table 9. Gender Classification in All ASD Subjects

Models	100 epochs		
	Training Accuracy	Test Accuracy	AUC
CNN Dropout: 0.5	0.9573	0.7804	0.8504
DNN Dropout: 0.5	0.9982	0.7238	0.7833

Number of scan-path images used from ASD and TD after augmentation are 2,118 and 1,767, respectively.

Bolded result indicate the best performance achieved on testing set.

Table 10. Gender/Age Classification in 5–9 Years Old Age Group of ASD Participants

Gender	DNN Augmented number	Dropout: 0.5		
		100 epochs		
		Training Accuracy	Test Accuracy	AUC
Male	3,121			
Female	2,905	0.9905	0.9884	0.9985
Male	1,276			
Female	1,259	0.9966	0.8581	0.9052
Male	961			
Female	957	1	0.8507	0.8194

The bold result indicate the best performance achieved.

are reported in Table 10 indicating 98% female and male stratification accuracy among 5 to 9 years old ASD population when the highest level of sample augmentation is used and the highest number of samples are generated. This performance is reduced when a lower level of sample augmentations are used but the results still indicate a substantial difference in response between female and male ASD participants.

Because there are few samples in other age groups, there is no further experiment on the gender classification of ASD patients in different age groups.

4.3 Spatio-Temporal Analysis I: The Importance of Temporal Information

The results presented in previous experiments attest to the following aspects:

- (1) Patterns of spatial information of eye gaze data captured in scan-path images are able to stratify ASD and TD participants (Table 3: DNN = 74.4%)
- (2) Scan-path patterns in children with ASD are influenced by Age (Table 5: DNN = 79.8%, CNN = 80.5% 2–4 years old children with ASD vs. 10–17 years old children with ASD)
- (3) Scan-path patterns in children with ASD are influenced by Gender (Table 9: DNN = 78.0% males ASD vs. female ASD)
- (4) Age and Gender factors together have dual impact scan-path patterns in children with ASD (Table 10: DNN = 98.8% male vs. female classification in 5–9 years old children with ASD)

The results are encouraging and indicative of some degree of success on mining the underlying spatial eye-gaze pattern differences between ASD and TD. The results with spatial eye-gaze scan-paths also speak to factors contributing to ASD heterogeneity e.g., Age and Gender. Spatial-based scan-paths capture the eye-gaze patterns by removing the temporal dimension of the data. These scan-path images, while providing an overall view of visited points on the screen, eliminate any pattern differences between ASD and TD in temporal dimension of the data. In order to better understand the importance of temporal dimension on eye-gaze scan-paths, a new set of analysis are performed.

Table 11. All Testing Results and Information of Five Periods for ASD and TD Classification Using CNN

Period	Trial	Train Acc	Test Acc	AUC	Used ASD No. of Images	Used TD No. of Images
0–3 S	1	100.00%	75.69%	81.03%	1,000	1,054
	2	100.00%	75.85%	81.11%		
	3	100.00%	77.47%	83.13%		
	AVE	100.00%	76.34%	81.76%		
3–6 S	1	99.93%	71.80%	78.55%	999	1,055
	2	100.00%	72.93%	79.84%		
	3	99.51%	70.50%	78.59%		
	AVE	99.81%	71.74%	78.99%		
6–9 S	1	99.52%	67.10%	76.61%	996	1,070
	2	100.00%	69.84%	75.65%		
	3	100.00%	70.97%	75.71%		
	AVE	99.84%	69.30%	75.99%		
9–12 S	1	100.00%	71.27%	76.48%	994	1,058
	2	100.00%	70.29%	77.23%		
	3	100.00%	71.43%	76.11%		
	AVE	100.00%	71.00%	76.61%		
12–15 S	1	100.00%	69.47%	74.77%	994	1,069
	2	100.00%	71.24%	76.51%		
	3	100.00%	69.95%	77.09%		
	AVE	100.00%	70.22%	76.12%		

The bold font highlights the best performing sub-window and the performance achieved by it in this temporal window.

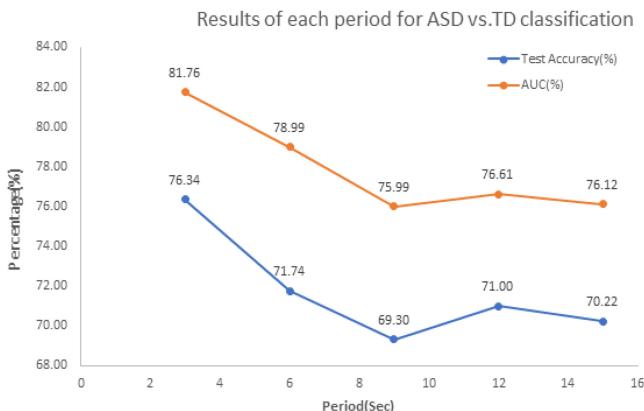


Fig. 6. Scatter plot of classification performances achieved by five temporal periods.

4.3.1 Stratification of ASD and TD Using Spatio-Temporal Eye-gaze Scan-paths. To understand ASD and TD pattern differences in temporal dimensions of eye-gaze scan-paths, non-overlapping 3 s temporal windows are considered and their associated scan-paths are assessed. The results are reported in Table 11. The evaluation is repeated 3 times and the outcome is averaged to represent final performance.

In order to understand the effect of each temporal period more intuitively, the average values of each period in Table 11 are used (see Figure 6). Since the training accuracies are always 100%, no curve is drawn separately.

Table 12. All Testing Results and Information of Five Periods for ASD and TD Classification

Video Name	Num. of times Viewed
sr04	90
AM_A4_S6_B4_GM_D1_F1	92
sr01	62
db0101b12	60

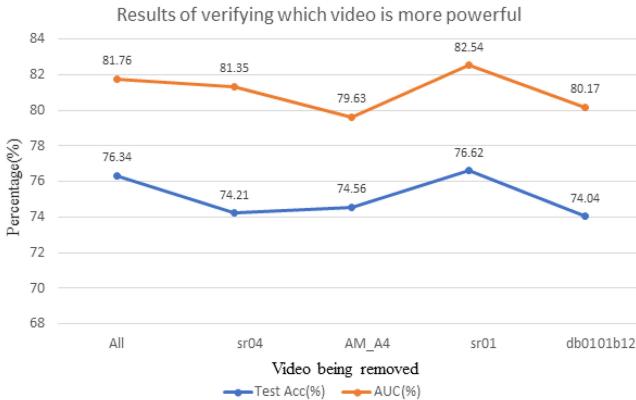


Fig. 7. Scatter plot of classification performances achieved by 0–3s temporal period after removing all scan-path samples of different video clips.

The results indicate that the first 3 s of video clips representing 0–3 s temporal window encapsulates the scan-path trajectories with highest ASD and TD stratification capability while the remaining periods are performing almost consistent. This is indicative that there are a degree of differences between ASD and TD eye-gaze patterns hidden in temporal dimensions of scan-path patterns.

4.3.2 Assessing the Performance Variations Impacted by the Video Clips Watched By Participants. Before having a closer look at the contribution of temporal information on stratifiability of scan-path patterns, it is necessary to consider the possible performance variation across video clips and their contribution to the observed performance.

Given the variations in number of times video clips were viewed by participants enrolled in this study, and considering the low number of views presented in Table 6 for some of the clips, only the four most viewed video clips are assessed in this analysis (See Table 12). Similar to previous experiments, in each evaluation, all scan-path samples generated from each of these four video clips are removed from dataset and the remaining samples are used for training and evaluation. The classification accuracy outcome of this experiment attests to the contribution of each video on stratification of ASD and TD. In this experiment, high and low classification accuracy achieved after removing samples of a given video clip is indicative of low or negative and high or positive contribution of such samples in stratification of ASD and TD, respectively.

Inspired by findings of previous experiments, only the 0–3 s tie window scan-paths are considered in this experiment. The evaluation with each video is repeated three times. The results are presented in Table 13.

In order to understand the effect of each video while only using scan-paths of 0–3 s temporal window, the average performances presented in Table 13 are used to draw the patterns presented in Figure 7.

Table 13. CNN Stratification Performance Achieved Using Scan-paths from 0–3 s Temporal Window After Removing Most Viewed Videos

Video Removed	Trial	Train Acc	Test Acc	AUC	Used ASD No. of Images	Used TD No. of Images
All Videos Included	1	100.00%	75.69%	81.03%	1,000	1,054
	2	100.00%	75.85%	81.11%		
	3	100.00%	77.47%	83.13%		
	AVE	100.00%	76.34%	81.76%		
sr04	1	100.00%	74.96%	80.78%	942	986
	2	100.00%	73.92%	81.48%		
	3	100.00%	73.75%	81.79%		
	AVE	100.00%	74.21%	81.35%		
AM_A4_S6_B4_GM_D1_F1	1	100.00%	73.75%	79.64%	944	985
	2	100.00%	76.17%	80.83%		
	3	100.00%	73.75%	78.43%		
	AVE	100.00%	74.56%	79.63%		
sr01	1	100.00%	76.34%	81.88%	952	1,034
	2	100.00%	77.18%	82.55%		
	3	100.00%	76.34%	83.20%		
	AVE	100.00%	76.62%	82.54%		
db0101b12	1	100.00%	72.30%	80.60%	953	1,019
	2	100.00%	73.14%	79.35%		
	3	100.00%	76.69%	80.56%		
	AVE	100.00%	74.04%	80.17%		

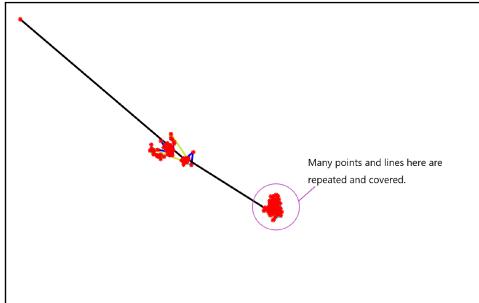


Fig. 8. Repeated points and lines in some cases.

The results reported in Table 13 and Figure 7 indicate that removing scan-path samples from first 3 s temporal window of sr01 video clip increases the overall stratification capability of the classifier. This attest to negative impact of scan-path samples generated from sr01 eye-gaze data.

The results also indicate omission of samples from sr04, AM_A4_S6_B4_GM_D1_F1, and db0101b12 causes the highest loss in ASD and TD stratification capability of the network attesting to the importance of these video clips in the observed overall performance.

It is noteworthy that in preparation of results presented in Tables 12 and 13 no augmentation is performed since the main purpose of the evaluation is to verify the impact of each video clip on overall performance.

4.4 Spatio-Temporal Analysis II: Digging Deeper in Presentation of Temporal Dimension

Looking at scan-path images, it is a common phenomenon to see multiple lines between close-by points. See Figure 8 as a zoomed example.

Table 14. First Color Setting Based on Velocity Range

Velocity Range	Color
0 (fixation)	Red Line
(0, 0.5]	Yellow Line
(0.5, 1]	Green Line
(1, 2]	Cyan line
(2, 3]	Blue line
(3, 5]	Mulberry Line
(5, 10]	Dark Yellow Line [R0.8, G0.9, B0.3]
(10, inf)	Black Line

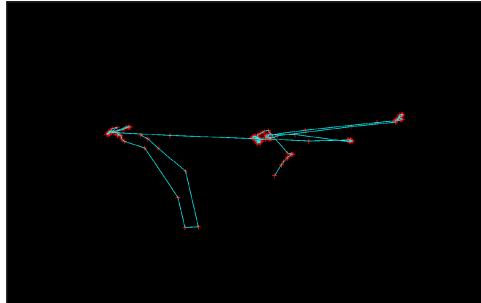


Fig. 9. Scan-path patterns of a subject without velocity information-based color-coding.

Aiming to increase the depth of information presented in scan-path representation of eye gaze data, factors such as average velocity between two points are considered using following equation

$$\text{Average Velocity } (P_1, P_2) = \frac{\text{Distance}(P_1, P_2)}{\text{Time Interval}(P_1, P_2)}, \quad (1)$$

where the distance between the two points is obtained by the Euclidean algorithm. P_1 and P_2 represent the x and y coordinates of two consecutive points. To incorporate both temporal and velocity information into scan-paths, the velocity values, generated from predefined temporal windows are used as color values of scan paths in the given time window. Table 14 represents the sets of velocity color mappings considered in this experiment. Another variation of this colormap where threshold values of the velocity are different is also considered and no considerable difference in the performance is observed.

An Exampled scan-path sample using these two color coding schemes is illustrated in Figures 9–11.

5 EXPERIMENTS AND RESULTS

ASD and TD Classification for Each Second in 3–6 sec Period: Aiming to deal with the unbalanced nature of samples in ASD and TD, the TD samples are augmented using rotation method to close the sample size gap between the two group. To obtain relatively accurate data, the test was repeated three times in each interval. In order to compare the test results with previous findings presented in Table 11, the test results of the 3 s to 6 s time window are added to the table. The results indicate a considerable increase in the overall performance of the 3–6 s tie window from 71.74% (see Table 11) to 80.7% (see Table 15). Table 16 provides a comparison between the use of velocity-based color-coded scan-paths and the original scan-paths. The 3–6 s time window results acquired by the velocity-based color-coded scan-paths is also performing considerably better than

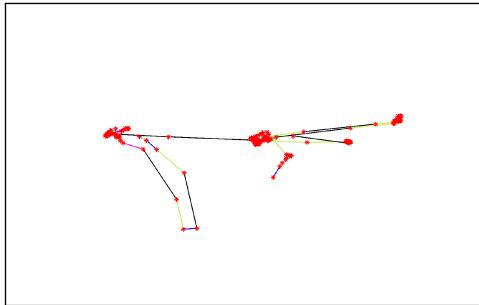


Fig. 10. Scan-path patterns of a subject using the velocity and the temporal information-based color-coding presented in Table 14.

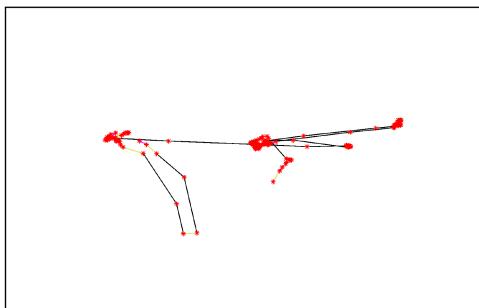


Fig. 11. Scan-path patterns of a subject using the temporal information-based color-coding presented in Table 14.

Table 15. ASD vs. TD Stratification Using Velocity-based Color Mapped Scan-paths in 3–6 sec Time Window Using Color Coding Scheme Presented in Table 14

Period	Trial	Train Acc	Test Acc	AUC	Used ASD No. of Images	Used TD No. of Images
3–4 Sec	1	100.00%	72.56%	78.63%	989	1,050
	2	100.00%	73.33%	80.37%		
	3	99.68%	72.74%	79.88%		
	AVE	99.89%	72.88%	79.63%		
4–5 Sec	1	99.98%	70.18%	77.93%	990	1,054
	2	100.00%	71.63%	77.48%		
	3	100.00%	71.95%	78.25%		
	AVE	99.99%	71.25%	77.89%		
5–6 Sec	1	99.62%	68.57%	76.03%	987	1,059
	2	100.00%	66.97%	74.92%		
	3	99.77%	67.55%	76.21%		
	AVE	99.80%	67.70%	75.72%		
3–6 Sec	1	100.00%	79.82%	84.33%	1,000	1,068
	2	100.00%	80.23%	83.89%		
	3	100.00%	80.69%	83.36%		
	AVE	100.00%	80.25%	83.86%		

the 1 s tie window intervals between the 3 s and 6 s period (see Table 15). A possible explanation for the observed increase in classification accuracy in 3–6 s time window compared to the 1s time intervals is the difference in the amount of information and patterns that can be captured by a

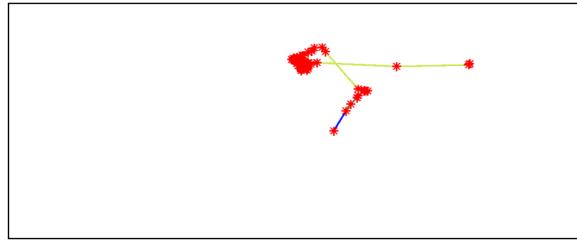


Fig. 12. Scan-path patterns of a subject using the velocity and color-coding presented in Table 14 in 3–4 s time window.

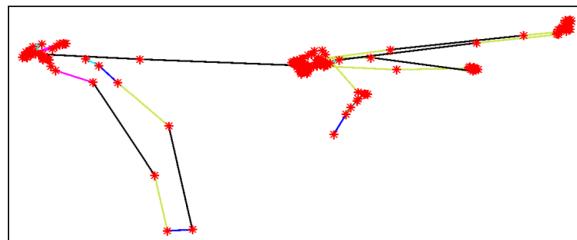


Fig. 13. Scan-path patterns of a subject using the velocity and color-coding presented in Table 14 in 3–6 s time window.

Table 16. Comparison of ASD vs. TD Stratification Performance Using Scan-paths with and Without Velocity-based Color Mapping Scan-paths in 3–6 s Time Window

Period	Trial	Train Acc	Test Acc	AUC	No. of images used ASD	No. of images used TD
Without velocity-based color mapping	1	99.93%	71.80%	78.55%	999	1,055
	2	100.00%	72.93%	79.84%		
	3	99.51%	70.50%	78.59%		
	AVE	99.81%	71.74%	78.99%		
With velocity-based color mapping	1	100.00%	79.82%	84.33%	1,000	1,068
	2	100.00%	80.23%	83.89%		
	3	100.00%	80.69%	83.36%		
	AVE	100.00%	80.25%	83.86%		

3 s time window compared to 1 s time window. Figures 12 and 13 provides an example of such scan-path differences between these two-time windows.

6 CONCLUSION

This study investigated the feasibility of velocity-infused spatiotemporal scan-path samples in stratifying ASD and TD diagnosis using DNN and CNN. Using eye gaze data from 99 children with and without ASD, it is shown that this approach is able to identify pattern differences associated with age, gender, and the mixed effect of gender and age. These are some of the well-known factors in autism heterogeneity that make it difficult to identify robust biomarkers for autism.

REFERENCES

- [1] R. Carette, F. Cilia, G. Dequen, J. Bosche, J. L. Guérin, and L. Vandromme. 2018. Automatic autism spectrum disorder detection thanks to eye-tracking and neural network-based approach. In *Proceedings of the International Conference on IoT Technologies for HealthCare*. Springer International Publishing, 75–81.
- [2] X. Huang, C. Shen, X. Boix, and Q. Zhao. 2015. SALICON: Reducing the semantic gap in saliency prediction by adapting deep neural networks. In *Proceedings of the IEEE International Conference on Computer Vision*. 262–270.
- [3] M. Kümmerer, L. Thesis, and M. Bethge. 2015. Deep gaze I: Boosting saliency prediction with feature maps trained on imagenet. In *Proc. Int. Conf. Learn. Represent. Workshops*.
- [4] M. Kümmerer, T. S. A. Wallis, and M. Bethge. 2016. Deepgaze ii: Reading fixations from deep features trained on object recognition. arXiv:1610.01563. Retrieved from <https://arxiv.org/abs/1610.01563>.
- [5] N. Liu, J. Han, D. Zhang, S. Wen, and T. Liu. 2015. Predicting eye fixations using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 362–370.
- [6] J. Pan, E. Sayrol, X. Giro-i Nieto, K. McGuinness, and N. E. O'Connor. 2016. Shallow and deep convolutional networks for saliency prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 598–606.
- [7] M. Jiang and Q. Zhao. 2017. Learning visual attention to identify people with autism spectrum disorder. In *Proceedings of the IEEE International Conference on Computer Vision*. 3287–3296.
- [8] M. Elbattah, R. Carette, G. Dequen, J. L. Guérin, and F. Cilia. 2019. Learning clusters in autism spectrum disorder: Image-based clustering of eye-tracking scanpaths with deep autoencoder. In *Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 1417–1420.
- [9] G. Pusiol, A. Esteva, S. S. Hall, M. Frank, A. Milstein, and L. Fei-Fei. 2016. Vision-based classification of developmental disorders using eye-movements. *Medical Image Computing and Computer-Assisted Intervention (MICCAI'16)*. 9901, 317–325.
- [10] Y. Tao and M. Shyu. 2019. *SP-ASDNet: CNN-LSTM based ASD classification model using observer ScanPaths*. In *Proceedings of the 2019 IEEE International Conference on Multimedia & Expo Workshops*. 641–646.
- [11] C. Wu, S. Liaqat, S. Cheung, C. Chuah, and S. Ozonoff. 2019. Predicting autism diagnosis using image with fixations and synthetic saccade patterns. In *Proceedings of the 2019 IEEE International Conference on Multimedia & Expo Workshops*. 647–650.
- [12] American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. Arlington, Va.: American Psychiatric Association, 2013.
- [13] G. Dawson, R. Bernier, and R. H. Ring. 2012. Social attention: A possible early indicator of efficacy in autism clinical trials. *Journal of Neurodevelopmental Disorders* 4, 1 (2012), 1–12.
- [14] C. Lord, M. Rutter, P. C. DiLavore, S. Risi, K. Gotham, and S. Bishop. 2012. Autism diagnostic observation schedule: ADOS-2. Western Psychological Services, USA.
- [15] K. Chawarska, S. Macari, and F. Shic. 2012. *Context modulates attention to social scenes in toddlers with autism*. *Journal of Child Psychology and Psychiatry* 53, 8 (2012), 903–913.
- [16] M. Rutter, A. LeCouteur, and C. Lord. 2003. Autism diagnostic interview-revised (ADI-R). Los Angeles, CA: Western Psychological Services
- [17] A. Klin, W. Jones, R. Schultz, F. Volkmar, and D. Cohen. 2002. Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry* 59, 9 (2002), 809.
- [18] C. D. Elliott. 2007. *Differential Abilities Scale II*. San Antonio, TX: Pearson Education, Inc.
- [19] K. Pierce, S. Marinero, R. Hazin, B. McKenna, C. C. Barnes, and A. Malige. 2016. Eye tracking reveals abnormal visual preference for geometric images as an early biomarker of an autism spectrum disorder subtype associated with increased symptom severity. *Biological Psychiatry*, Elsevier, 79, 8 (2016), 657–666. DOI: <https://doi.org/10/f8drwn>
- [20] C. Karatekin. 2007. Eye tracking studies of normative and atypical development. *Developmental Review* 27, 3 (2007), 283–348.
- [21] H. Kopka and P. W. Daly. 1999. *A Guide to L^TE_X*, 3rd ed. Harlow, England: Addison-Wesley.
- [22] C. Romuald, M. Elbattah, G. Dequen, J. L. Guérin, and F. Cilia. 2018. Visualization of eye-tracking patterns in autism spectrum disorder: Method and dataset. In *Proceedings of the 2018 13th International Conference on Digital Information Management*. IEEE, 248–253.
- [23] C. Romuald, M. Elbattah, G. Dequen, J. L. Guérin, F. Cilia, and J. Bosche. 2019. Learning to predict autism spectrum disorder based on the visual patterns of eye-tracking scanpaths. In *Proceedings of the 12th International Conference on Health Informatics*.
- [24] M. Elbattah, C. Romuald Carette, D. Gilles, J. L. Guérin, and F. Cilia. 2019. Learning clusters in autism spectrum disorder: Image-based clustering of eye-tracking scanpaths with deep autoencoder. In *Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 1417–1420.

- [25] A. Masi, M. N. DeMayo, M. Glozier, and A. J. Guastella. 2017. An overview of autism spectrum disorder, heterogeneity and treatment options. *Neuroscience Bulletin* 33, 2 (2017), 183–193.
- [26] E. Fombonne. 2009. Epidemiology of pervasive developmental disorders. *Pediatric Research* 65, 6 (2009), 591–598.
- [27] D. M. Werling and D. H. Geschwind. 2013. Sex differences in autism spectrum disorders. *Current Opinion in Neurology* 26, 2 (2013), 146–153.
- [28] D. H. Geschwind and P. Levitt. 2007. Autism spectrum disorders: Developmental disconnection syndromes. *Current Opinion in Neurology* 17, 1 (2007), 103–111.
- [29] R. Rizzo and P. Pavone. 2016. Aripiprazole for the treatment of irritability and aggression in children and adolescents affected by autism spectrum disorders. *Expert Rev Neurother* 16, 8 (2016), 867–874.
- [30] F. Shic, G. Chen, M. Perlmuter, E. Gisin, A. Dowd, E. Prince, L. Flink, S. Lansiquot, C. Wall, E. Kim, Q. Wang, S. Macari, and K. Chawarska. 2014. Components of limited activity monitoring in toddlers and children with ASD. In *Proceedings of the 2014 International Meeting for Autism Research*.
- [31] F. Shic, J. Bradshaw, A. Klin, B. Scassellati, B., and K. Chawarska. 2011. Limited activity monitoring in toddlers with autism spectrum disorder. *Brain Research* 1380 (2011), 246–254.
- [32] K. Chawarska, S. Macari, and F. Shic. 2012. Context modulates attention to social scenes in toddlers with autism. *Journal of Child Psychology and Psychiatry* 53, 8 (2012), 903–913.
- [33] K. Chawarska, S. Macari, and F. Shic. 2013. Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorders. *Biological Psychiatry* 74, 3 (2013), 195–203.
- [34] C. Karatekin. 2007. Eye tracking studies of normative and atypical development. *Developmental Review* 27, 3 (2007), 283–348.
- [35] W. Jones, K. Carr, and A. Klin. 2008. Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. *Archives of General Psychiatry* 65, 8 (2008), 946–954.
- [36] D. J. Campbell, F. Shic, S. Macari, and K. Chawarska. 2013. Gaze response to dyadic bids at 2 years related to outcomes at 3 years in autism spectrum disorders: A subtyping analysis. *Journal of Autism and Developmental Disorders*. 44, 2 (2014), 431–442.
- [37] S. Feinman. 1982. Social referencing in infancy. *Merrill-Palmer Quarterly* 28, 4 (1982), 445–470.
- [38] T. A. Walden and T. A. Ogan. 1988. The development of social referencing. *Child Development* 59, 5 (1988), 1230–1240.
- [39] G. Dawson, K. Toth, R. Abbott, J. Osterling, J. Munson, A. Estes, and J. Liaw. 2004. Early social attention impairments in autism: Social orienting, joint attention, and attention to distress. *Developmental Psychology* 40, 2 (2004), 271–282.
- [40] A. Senju, V. Southgate, S. White, and U. Frith. 2009. Mindblind eyes: An absence of spontaneous theory of mind in asperger syndrome. *Science*, 325, 5942 (2009), 883–885.
- [41] B. Li, E. Barney, C. Hudac, N. Nuechterlein, P. Ventola, L. Shapiro, and F. Shic. 2020. Selection of eye-tracking stimuli for prediction by sparsely grouped input variables for neural networks: Towards biomarker refinement for autism. In *Proceedings of the ACM Symposium on Eye Tracking Research and Applications, Stuttgart*. 1–8.
- [42] B. Li, N. Nuechterlein, E. Barney, C. Foster, M. Kim, M. Mahony, A. Atyabi, L. Feng, Q. Wang, P. Ventola, L. Shapiro, and F. Shic. 2021. Learning oculomotor behaviors from scanpath. In *Proceedings of the 23rd ACM International Conference on Multimodal Interaction*. 1–13.
- [43] F. Shic, A. J. Naples, E. C. Barney, S. A. Chang, B. Li, T. McAllister, M. Kim, K. J. Dommer, S. Hasselmo, A. Atyabi, Q. Wang, G. Helleman, A. R. Levin, H. Seow, R. Bernier, K. Charwaska, G. Dawson, J. Dziura, S. Faja, S. S. Jeste, S. P. Johnson, M. Murias, C. A. Nelson, M. Sabatos-DeVito, D. Senturk, C. A. Sugar, S. J. Webb, and J. C. McPartland. 2022. The autism biomarkers consortium for clinical trials: Evaluation of a battery of candidate eye-tracking biomarkers for use in autism clinical trials. *Molecular Autism* 13, 1 (2022), 15.

Received 15 July 2021; accepted 8 May 2022